Innovation vs Safety: A critical examination of regulatory approaches to artificial intelligence

Jennifer Graham

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

Submitted for Examination December 2022

Contents

Abstract

The functioning of society has been forever changed by the creation of artificial intelligence (AI); what was once a futuristic and unrealistic invention of science fiction now pervades our everyday lives in inconceivable ways in certain parts of the world. In the UK, we are almost always in direct contact with an AI-based system, whether this be via the smart phones we keep in our pockets, when competing our weekly shopping in-store or online, or when we are deciding what to watch at the end of a busy day. The ubiquitous state of this technology gives rise to both curiosity and concern. Whilst we may acknowledge that this technology streamlines and simplifies several of our daily activities and tasks, it is without question that there is an underlying unease associated with AI use due to its capricious nature.

It is clear that efforts to develop increasingly sophisticated AI and implement these systems at any given opportunity will not cease, therefore we must give justifiable consideration to the regulation of this technology, which at present is considerably sparse. This thesis therefore proposes an innovative approach to AI regulation, and in doing so examines in detail the various risks associated with AI use (using bias and discrimination within AI systems as a case study) and scrutinises the regulatory proposals for governing this technology presented by a variety of national, regional, and international bodies and states, in order to assess the current state of global readiness for AI governance. Overall, this thesis purports that there is a significant lack of research in reasonable and workable governance measures for AI, and as such considerable work must be undertaken in this space.

To this effect, a predominantly doctrinal and comparative approach is taken to this interdisciplinary project, and the research undertaken within this thesis contributes to the literature in several ways. Firstly, the thesis presents an internationally comparative analysis of AI regulatory proposals. This analysis examines, in depth, the various weaknesses of these regimes and proposes reasonable amendments to such. These proposals are made with a view to being robust, realistic and workable, and therefore have the capacity to be truly impactful. Secondly, this thesis features a detailed evaluation of the issues and key features necessary within any regulatory regime specifically targeted at governing modern technologies. An examination of this kind is lacking in current scholarship in this area, and so the work undertaken in this thesis contributes to this gap in the literature. Finally, by presenting reasonable recommendations, a workable proposed framework for harmonisation, including a rights-based impact assessment unique to this work, this thesis makes another original contribution to research in this space. This contribution is informed by the findings in the initial chapters of this thesis and encourages policymakers and professionals in this space to think innovatively about how we regulate AI.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of any other university or other institute of learning. Full acknowledgement has been given to all sources used.

Chapter Two of this thesis has been published as a standalone chapter in A. Lui, N. Ryder, (eds) FinTech, Artificial Intelligence and the Law: Regulation and Crime Prevention (Routledge, 2021). The Chapter has been edited for inclusion within this thesis.

Acknowledgements

I would like to firstly express my deepest gratitude to my Director of Studies Dr. Alison Lui, without whom I would have never happened upon this fascinating field of research. Thank you for sparking my interest in AI during my Masters degree and introducing me to this incredible field that has now become such a huge part of my life. Thank you for your endless guidance, encouragement and confidence in my abilities, and for not only helping me to complete my PhD journey but for also giving me a start in teaching – I am eternally grateful.

I would also like to wholeheartedly thank the staff at Liverpool John Moores University, specifically those within the School of Law, for your support and teachings throughout my entire University career. You have supported me consistently for almost a decade; from the day I entered as a first-year law student, right the way through to my final year of PhD studies.

Finally, this thesis would not have been possible without the boundless love and support of my dear friends and family, particularly that of my Mum, Dad and Daniel, and of course my little Oscar. Thank you for your abundant and unwavering support throughout this journey, for always being patient with me and for helping me to remain positive throughout – I will never be able to thank you enough. And my friends; to Sarah, Vikki, Rebecca, Lisa, Lauren, Eve and Elizabeth, thank you for bearing with me throughout my (lengthy) studies and always being my cheerleaders, for celebrating all my little achievements along the way and always encouraging me to persevere.

List of Figures

Table of Legal Instruments

African Commission on Human and Peoples' Rights, 'Resolution on the need to undertake a Study on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa' - ACHPR/Res. 473 (EXT.OS/ XXXI) 2021

American Convention on Human Rights 1969

AL 2021-344 SB78

Charter of Fundamental Rights of the European Union (2000/C 364/01)

Charter of Fundamental Rights of the European Union 2012/C 326/02

CO 2021 S.B. 169

Data Protection Act 2018

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. OJ L 281, 23.11.1995

Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA

European Commission, 'Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) COM(2022) 496 final

E.O. 13859 2019 84 FR 3967

European Commission, 'Proposal for a Regulation of the European Union and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts' COM (2021) 206 final

European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM (2020) 825 final)

European Commission, Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) (COM (2020) 842 final)

European Commission, Proposal for a Regulation of the European Parliament and of the Council on machinery products (Machinery Regulation) (COM (2021) 202 final)

European Commission, Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) (COM (2020) 767 final)

European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950

HI H.B. 454

Human Rights Act 1998

IL 2021 H.B. 53

IL 2021 H.B. 645

Loomis v Wisconsin 881 N.W.2d 749 (2016)

Loomis v Wisconsin 881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S.Ct. 2290 (2017)

MA H.B. 136

MA H.B. 142

MA S. 60

MS 2021 H.B. 633

Product Security and Telecommunications Infrastructure Bill, HL Bill 16 58/3

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

The Law on the Protection of Personal Data ('the Data Protection Law') issued under Resolution No.151 of 2020

United Nations Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects as amended on 21 December 2001

Universal Declaration of Human Rights 1948, 217 A (III)

WA 2021 S.B. 5092

List of Abbreviations

AI – Artificial Intelligence

AU – African Union

CAC – Cyberspace Administration of China

COMPAS – Correctional Offender Management Profiling for Alternative Sanctions

DCMS – Department for Digital, Culture, Media and Sport (UK)

DTSA – Digital Transformation Strategy for Africa

DfE – Department for Education

EU – European Union

FDA – Food and Drug Administration

FTC – Federal Trade Commission

GDP – Gross Domestic Product

IoT – Internet of Things

NAIAC – National Artificial Intelligence Advisory Committee (US)

NCSC – National Cyber Security Centre

NGAIDP – New Generation Artificial Intelligence Development Plan (China)

NHTSA – National Highway Traffic Safety Administration (US)

NIST – National Institute of Standards and Technology (US)

OECD – Organisation for Economic Co-Operation and Development

PSTI - Product Security and Telecommunications Infrastructure Bill

SAE – Society of Automotive Engineers

SDO – Standards Development Organisation

SME – Small and medium sized enterprises

STEM – Science, Technology, Engineering and Mathematics

UK – United Kingdom

UN – United Nations

UNESCO - United Nations Educational, Scientific and Cultural Organization

US – United States of America

Introduction

This thesis examines the ways in which we use artificial intelligence, and the need for regulation in this space. This examination includes an evaluation of some of the ethical and legal impacts, and subsequent implications, resulting from an increased use of AI. Following an examination of varying approaches to AI regulation observed across a variety of jurisdictions, this thesis discusses a number of constructive ways in which we could regulate the use of artificial intelligence in order for us to safely reap its benefits. This is essential; legal measures that are too restrictive will negatively impact the development of AI, and stifle innovation, which is now a pivotal part of most modern economies. Meanwhile, this will allow for a legal vacuum to grow in this space, which will ultimately be detrimental to all members of society. As a result, when regulating we must ensure that the interests of those designing AI are balanced with the wellbeing of society at large, ensuring the safety of the people.

Analysis of the literature suggests that there is confusion and disagreement as to the most effective regulatory methods to be used in the context of artificial intelligence; but it is held that in order to prevent this technology from being used improperly, an adequate and functional regulatory system should exist. This thesis therefore proposes a new regulatory framework that aims to enable international harmonisation in the face of AI regulation. A framework that incorporates aspects of statutory regulation and self-regulation, in order to advocate for AI safety, whilst still promoting innovation. To achieve this, several key research themes and questions were identified, and predominantly doctrinal and comparative research methods were utilised in order to provide the basis for a robust and rigorous piece. The following sections detail these research themes, the methodologies considered for this thesis and the overall structure of the work.

Key Research Themes

During the course of reviewing the literature for this thesis, several themes were identified which will form the main focus of this research; these themes include the current state of AI regulation, the widespread and relatively unchecked use of AI, and the occurrence of unintended consequences as a result of this AI use. For each of the themes, the relevant literature will be briefly assessed here.

The Current State of AI Regulation

Unsurprisingly, the majority of recent literature on emerging technologies focuses on the use of such by large organisations and institutions in order to target individuals in the course of

political campaigns or for commercial purposes for example.[1] Whilst this is of significance, and research in the area is vital, the literature on the whole appears to focus on the impacts of the use of artificial intelligence (AI) and less so on how we could be regulating AI and preventing these incidents from occurring.

Some research does detail the current state of regulation in relation to AI, however, this research typically examines this regulation without providing any constructive solutions. Sullivan's paper on the relationship between AI and GDPR illustrates this.[2] Whilst this research is valid, informative and interesting, it serves to highlight the disparity between the use of AI (usually by large organisations) and the rights that GDPR exists to protect; yet it fails to suggest constructive ways in which this gap could be bridged.

Likewise, there are researchers who are considering the idea of accountability relating to algorithm-based decision making; focusing on the prospect of using liability in order to regulate AI.[3] Despite taking a step in the right direction, this approach is interpreted as being very narrow in its application and would only apply to certain uses of algorithms, and as such would require further development before becoming a reasonable and workable solution in regulating AI.

Katyal's paper on accountability in relation to AI is probably the closest aligned to the recommendations of this thesis, in that Katyal considers the potential of self-regulation within large organisations in order to combat the misuse of data and technology.[4] Again, whilst this research is valuable and makes valid suggestions for the regulation of AI, it puts the onus for regulations on organisations and institutions, as the role of the government and national level regulation is discounted.[5] This appears to be a relatively common trend within the literature, therefore there appears to be a gap in research that this thesis aims to contribute to, by providing suggestions for effective and functional ways in which we can realistically regulate AI, resulting in solutions that are practicable.

---

[1] H. C. Boyte, 'John Dewey and Citizen Politics: How Democracy Can Survive Artificial Intelligence and the Credo of Efficiency' (2017) *Purdue University Press Education & Culture* 33(2) 13-48; R. M Glassman, 'Will Artificial Intelligence (AI) Make Democracy Irrelevant?' in *The Future of Democracy* (Springer, 2019) 189-198; A. K. Cybenko, G. Cybenko, 'AI and Fake News' (2018) *IEEE Intelligent Systems* 33(5) 1-5 *2*

[2] C. Sullivan, 'GDPR Regulation of AI and Deep Learning in the Context of IoT Data Processing – A Risky Strategy' (2018) *Journal of Internet Law* 22(6) 1-18 *8*

[3] S. C. Olhede, P. J. Wolfe, 'The growing ubiquity of algorithms in society: implications, impacts and innovations' (2018) *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376(2128) <https://doi.org/10.1098/rsta.2017.0364> accessed 20/11/2022

[4] S. K. Katyal, 'Private Accountability in the Age of Artificial Intelligence' (2019) UCLA Law Review 66(1) 54-142 *108*

[5] ibid

The literature pertaining to the current state of regulation of AI is informative and valuable, but lacking in certain aspects, most notably in the area of reforming existing regulation and the creation of a new regime. Some research begins to consider the potential for reform but falls short by making the case for regulation that has several flaws[6], or regulation that excludes prominent sectors[7], and as a result it appears that few have considered formulating a type of regulation that operates across sectors and is inclusive of different models and uses of AI, which is the aim of this thesis.

Following recent proposals by the European Union (EU) to create an AI Act to regulate the technology via a blanket style governance framework,[8] we have seen an increase in literature on the topic.[9] Again though, we see a recurring theme in that the shortcomings of these proposals are discussed, yet little is offered in terms of meaningful suggestions for amendments.

As a result, Chapters Three and Four of this thesis examine and compare the regulatory approaches proposed by a variety of nations, regions, and international bodies. In doing so, this thesis goes further by using this critical comparison to make suggestions for reasonable improvements to these existing and proposed regimes, and also uses this as grounds for a proposed new regulatory model in Chapter Six.

The widespread and unchecked use of AI

Regarding the use of AI, the literature confirms that little is actually understood about the outcome of organisations making use of AI in order to process and better understand the plethora of data we now have available to us, and that further investigation is warranted.[10] Even from a simple literature search, it is evident how widely big data is actually used across sectors with research taking place on the use of big data in healthcare[11], in business[12] and in education[13], to name just a few examples. With this in mind and taking into consideration the

[6] ibid n3
[7] ibid n4
[8] Proposal for a REGULATION OF THE EUROPEAN PARLIAMTN AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS COM/2021/206 final
[9] C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, L. Floridi, 'Artificial Intelligence and the 'Good Society': the US, EU, and UK approach' (2018) *Science and Engineering Ethics* 24 505-528
[10] C. Fredriksson, 'Big data creating new knowledge as support in decision-making: practical examples of big data use and consequences of using big data as decision support' (2018) *Journal of Decision Systems* 27(1) 1-19 *1*
[11] M. Cohn Adulamy, V. Shalev, 'Colonscore: The Use of Machine Learning of Big Data to Detect Colorectal Cancer' (2018) *Harefuah (Israel Medical Association)* 157(10) *634*
[12] R. Glass, S. Callahan, *The big-data driven business: how to use big data to win customers, beat competitors, and boost profits* (Wiley, New Jersey 2015)
[13] C. Matthew, D. Halliday, 'Big Data and the Liberal Conception of Education' (2017) *Theory and Research in Education* 15(3) 290-306 *295*

popularity of utilising big data, the point made by Fredriksson in that we know very little about the impact of the synthesisation of big data by AI[14], does become concerning.

The figures themselves display clearly just how astronomical the growth of data is and how it is becoming the most popular and valuable asset for many. In research conducted by Helbing et al, it was disclosed that in around 10 years' time there will be around 150 billion 'networked measuring sensors' (these are methods by which data is collected such as Google searches or Tweets), and these sensors will collect data that will double in total every 12 hours.[15] Thus, the literature provides a clear display of the extent to which data is growing, and a succinct narrative on how this data is being used and synthesised.

Despite this however, it appears as though the literature that discusses the uses of big data focuses on just that, the uses, and fails to fully examine any potential ways that we can safely manage and regulate the synthesisation of big data by AI.[16] This is evident in that a number of scholars seem to focus on the visualisation of data, and how technology can be improved in order to utilise data more efficiently.[17] Although this research is worthwhile, again, it appears to leave a gap in which the scope of this thesis falls, and that is to explore in detail the ethical implications of widespread use of AI and suggest ways in which we can manage its use and avoid legal implications.

In addition to this, it is interesting to note that much of the literature in this area focuses on the impact that using big data will have upon democracy primarily in the United States of America.[18] Once again, whilst papers such as that authored by Bertot examine and identify the key issues that are posed by the use of big data in relation to democracy and constitutional law, it fails to fully examine ways in which the issue could be tackled.

This appears to be an established trend in the literature, and clearly demonstrates the void within existing research that thesis aims to fill.

Unintended consequences as a result of AI use

The occurrence of unintended consequences as a result of AI use (specifically algorithmic bias and discrimination), is another theme that is focused on in detail during the course of

[14] ibid n10
[15] D. Helbing, B. S. Frey, G. Gigerenzer, E. Hafen, M. Hagner, Y. Hoffstetter, J. Van den Hoven, R. V. Zicarij, A. Zwitter, 'Will Democracy Survive Big Data and Artificial Intelligence?' in D. Helbing (eds) *Towards Digital Enlightenment* (Springer, 2018) 73-98 *80*
[16] ibid
[17] P. Simon, *The visual organization: data visualization, big data and the quest for better decisions* (Wiley, New Jersey 2014) 1
[18] J. C. Bertot, 'Social Media, Open Platforms, and Democracy: Transparency Enabler, Slayer of Democracy, Both?' *Proceedings of the 52nd Hawaii International Conference on System Sciences 2019* https://scholarspace.manoa.hawaii.edu/bitstream/10125/61631/0782.pdf accessed 22/11/2022

the thesis. As a result, the literature was consulted to find comprehensive examples of some of the most notable consequences arising from AI use.

Huq makes a very interesting observation in that the amount of literature relating to the concept of algorithmic unfairness and discrimination has actually generated such a large volume of definitions within the literature, that it is difficult to find one to describe the concepts succinctly.[19] This statement in itself displays how vast the literature is on data bias; however, this literature lacks meaningful solution for attending to the issue by means of regulation and governance.

Baeza-Yates actually states in his research that in order to eliminate or lessen the impact of algorithmic bias we need to make ourselves aware of our own biases, which is a notion supported by this thesis.[20] Despite this, it would appear that again, the literature does not go far enough in suggesting ways in which we could regulate in order to prevent or lessen the impacts of algorithmic bias.

Huq's article for example is a very comprehensive and formative piece on the matter of algorithmic bias but only outlines the key issues, identifying that current laws are not sufficient to tackle the issue.[21] Whilst this is useful research, it falls short in suggesting realistic ways to tackle the issue, and therefore provides grounds for research such as that contained in this thesis, that attempts to go further in making those suggestions (see Chapter Two).

One commonality can be established throughout the literature however; the majority of the literature on AI-based legal concerns discusses solutions in the short term.[22] As a result, there is an element of contradiction and uncertainty within the literature regarding how AI should be regulated, and there seems to be agreement that there is no one effective model to be followed.[23] This is likely due to the nature of AI itself; it is a relatively new subject to be studied from the legal perspective.[24] Therefore, the literature presents an opportunity for research such as that presented within this thesis to have meaningful impact.

[19] A. Z. Huq, 'Racial Equality in Algorithmic Criminal Justice' (2019) *Duke Law Journal* 68(6) 1043-1134 *1115*
[20] Baeza-Yates, R. 'Bias on the Web' (2018) *Communications of the ACM* 61(6) 54-61 *55*
[21] ibid n19
[22] G. Merchant, '"Soft Law" Governance of Artificial Intelligence. UCLA: The Program on Understanding Law, Science and Evidence (PULSE)' (escholarship.org, 2019) https://escholarship.org/uc/item/0jq252ks accessed 22/11/2022
[23] ibid
[24] ibid

Research Questions

As a result of this literature review and identification of the key research themes, several research questions were identified that will guide the work undertaken in this thesis. They are as follows:

1. What are the most pertinent issues and threats posed by AI?
2. How well equipped are current legal instruments, proposed legal instruments, strategies and frameworks in dealing with the issues posed by AI?
3. What key regulatory principles are valuable and should be included in an ideal AI governance framework?
4. What realistic and workable recommendations can be made to improve the current state of AI regulation?

Thesis outline and contribution to research

This thesis contains six substantive chapters, as well as an introduction and a conclusion. Chapter One addresses the relationship that exists between artificial intelligence and the law; this is a contextual chapter that introduces AI, relevant definitions and refers to some established issues stemming from the use of AI in various capacities and sectors. This Chapter is foundational and provides necessary context for the substantive work to follow. This Chapter begins to answer the first research question, namely what are the most pertinent issues and threats posed by AI?

The approach taken within this thesis and the proposals made within allow for a number of original contributions to be made to the scholarship on the topic of AI regulation. Chapter Two provides a unique look at AI-based bias and discrimination as an in-depth case study. This Chapter was published in July 2021 in 'FinTech, Artificial Intelligence and the Law: Regulation and Crime Prevention', edited by Alison Lui and Nicholas Ryder and has been amended for inclusion within this thesis. This Chapter provides reasonable suggestions for the amendment of current legislative regimes such as GDPR[25] and the Equality Act 2010.[26] This Chapter deals with the first and second research questions of this thesis; what are the most pertinent issues and threats posed by AI (a specific focus is given to the discrimination and bias problem here), and how well equipped are current legal instruments in dealing with the issues posed by AI?

[25] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1
[26] Equality Act 2010

Chapters Three and Four offer an in-depth critical evaluation and comparison of approaches to AI regulation taken by a selection of nations, regions and international organisations. This comparison of AI regulatory approaches is unique, as the literature at present lacks such a widespread analysis. Both chapters are particularly useful in forming a foundation for the proposals to follow later in the thesis. This comparative analysis is structured in such a way that it considers both the successes and shortcomings of these proposed legal frameworks and AI strategies, and suggests reasonable amendments to improve these approaches where appropriate. The key difference between legal frameworks and strategies here being that the legal frameworks are proposed legislative instruments intended to have legally binding authority, whereby strategies are documents released by governments and other bodies that aim to signal their intention for AI regulation and innovation planning.

Therefore, these chapters feature analysis of a variety of legal frameworks, strategies and AI initiatives, all of which have differing levels of legally binding authority. As a result, these chapters primarily tackle the second research question posed by this thesis; how well equipped are current legal instruments, proposed legal instruments and strategies in dealing with the issues posed by AI?

Chapter Five follows suit by contributing an innovative look at the issues present when regulating modern technology; it does this by drawing upon recognised tech-related regulatory issues, key principles necessary to any AI regulatory regime, and by evaluating current approaches to regulating different types of technology such as IoT (Internet of Things). This Chapter therefore answers the third research question presented in this thesis; what key regulatory principles are valuable and should be included in an ideal AI governance framework?

These Chapters provide a ground for Chapter Six, which contains a comprehensive proposal for a new regulatory method for governing AI, which comes in the form of a rights-based impact assessment. The proposal contained within this Chapter is an original contribution to the field of AI regulation and is highly implementable, therefore resulting in research that is high impact in nature. This final chapter therefore deals with the final research question presented by this thesis; what realistic and workable recommendations can be made to improve and secure the current state of AI regulation?

Methodology

Introduction

"All methodologies have strengths and limitations, and each brings with it ethical values and principles on how to conduct research".[27] Choosing the most suitable and beneficial research methods for one's research is an important decision, and one that will have considerable influence over the final resulting work. The methodology is the connection that bridges the gap between the initial research question and the overall findings and results, which makes the theoretical framework one of the most important parts of the thesis.[28] One must therefore be aware of the notable criticisms and issues associated with a particular research method in order to make more informed choices regarding their use.[29]

There has long been debate regarding the most effective research methods to use within law, and considerable discussion regarding how traditional research methods that are common within other fields should apply to legal research.[30] In part, this is likely due to the emergence of new traditions in legal scholarship, such as the evolution of studying 'law in context' as opposed to the more traditional study of black letter law.[31] The development of other areas of law, such as the law relating to AI has confirmed the fact that new legal traditions are emerging and so our choice of legal research methods must also adapt to accommodate these new areas. This does not mean we must select entirely new methodologies and forget traditions, but that we must be innovative with the way in which we approach this new era of legal research.

As we see these relatively new legal traditions develop, there is also inevitably going to be some confusion regarding the overlap of subject areas and the best way to approach researching these new interdisciplinary topics (such as AI and law). Later in this section, the role of interdisciplinary research within this thesis will be explored in some detail, but rather interestingly we are seeing scholars such as Al Amaren et al begin to refer to law as a

---

[27] L. Blair, 'Choosing a Methodology' in *Writing a Graduate Thesis or Dissertation* (Sense Publishers, Rotterdam 2016) *49*

[28] E. M. Al Amaren, A. M. A. Hamad, O. F. Al Mashhour, M. I. Al Mashni, 'An Introduction to the Legal Research Method: To Clear the Blurred Image on How Students Understand the Method of Legal Science Research' (2020) *International Journal of Multidisciplinary Sciences and Advanced Technology* 1(9) 50-55 *51*

[29] Ibid n27

[30] M. Salehijam, 'The Value of Sysetmic Content Analysis in Legal Research' (2018) *Tilburg International Law Review* 23(1-2) *34*

[31] M. McConville, W. H. Chui, *Research Methods in Law* (Edinburgh University Press, Edinburgh 2007). 'Law in context' refers to the study of law as a potential driver of societal issues as opposed to always remaining a solution and acknowledges the benefits of other non-law solutions such as political re-arrangements.

'science' for the purposes of methodology selection, perhaps in response to the diverse legal research landscape we are now seeing.[32]

Choosing the most effective research methods for one's study is therefore more interesting than ever, with legal research becoming ever more expansive, a wider variety of methodologies are available to choose from. The following sections consider various research methods that underpin this thesis alongside their potential limitations, and how these have been mitigated.

Types of research method

The study of law, and research conducted in this space, is done so with a view to achieve a better understanding of the law or legal problem, and to make contribution to the law typically in an effort to minimise an existing legal problem. Therefore, most legal research to some degree will focus on the analysis of existing regulation and case law, or in the case of this thesis, on regulatory proposals where existing regulation is lacking. Because of this, there are certain methodologies that lend themselves more closely to this type of research activity, namely a critical doctrinal methodology and comparative methodology. However, several other methodologies also underpin the research conducted during this thesis and are discussed in the following sections.

Qualitative methods in law

Legal researchers will often proclaim that their most used and effective research method is a critical doctrinal one. Whilst this might be true, the doctrinal approach forms part of a wider methodology used across most of the social sciences and humanities subjects. As Webley states, even though we usually associate some research methods such as qualitative methodologies with fields other than law, law students and legal professionals are using qualitative methods every day, despite not necessarily recognising that they are.[33] This was a particularly interesting finding, and one that has shaped how I approached the methodological choices for this thesis; I wanted to better understand how the various methodologies chosen for this thesis interlinked, and also if they were actually part of a more broad research method without my realising.

Qualitative research can be described as "watching people in their own territory" and being "naturalistic and participatory".[34] Whilst this definition might not align directly with the research undertaken within this thesis, this research does aim to 'understand and solve a

[32] Ibid n28
[33] L. Webley, 'Qualitative Approaches to Empirical Legal Research' in P. Cane and H. Kritzer (eds) Oxford Handbook of Empirical Legal Research (Oxford University Press, 2010)
[34] J. Kirk, M. L. Miller, 'Reliability and Validity in Qualitative Research' (SAGE Publications, 1986)

problem from a humanistic approach', which Pathak et al determine is indicative of a qualitative study.[35] In addition to these definitions of qualitative research methods, "Qualitative research is particularly good for examining whether or not a particular social phenomenon exists and if so, the nature of the phenomenon".[36] Establishing the existence of several ethical and legal issues associated with AI, such as accountability, transparency and non-discrimination, was necessary in order to make the case for regulation in this space, therefore it would seem that the methods used thus far within this thesis fit within this qualitative category.

Further to this, classical content analysis is a method used to examine existing texts and other documents and sits at the border of qualitative and quantitative methods, and strikes similar resemblance to the doctrinal method used by most legal scholars.[37] It is arguable therefore that the traditional critical doctrinal methodology typically used when studying the black letter law, and even the comparative methodology employed within this thesis, sit within this classic category of qualitative legal research.

However, there are some common hallmarks of the qualitative methodology that are not present within this thesis for several reasons. Initially, within the planning stages of this thesis, it had been intended that interviews with professionals within this space might be carried out in order to add to and enhance the existing findings. This may have taken the form of either in person interviews or questionnaires. However, the decision was made to avoid using these methods, and instead use those already discussed within this section.

This choice was made for several reasons; firstly, much of the content of this thesis focuses specifically on the analysis of regulatory proposals, and as such analysing these documents themselves yields much more insight than considering an individual's thoughts on the flaws and benefits of these proposals. Not to mention that most specialists in this field have been involved either in the formation of these proposals and/or have been involved in the formal commenting and amendment process. Therefore, the findings of these interviews and questionnaires would not have added any significant contribution to the literature already available on this subject and would not further the findings of this thesis.

Secondly, there are several strategic, ethical, and legal issues associated with conducting either non-structured interviews, focus groups or even using questionnaires. Questionnaires would need to contain closed questions, and it would have been difficult to turn the findings

[35] V. Pathak, B. Jena, S. Kalra, 'Qualitative research' (2013) *Perspectives in Clinical Research* 4(3) 192
[36] Ibid n33
[37] Ibid n33

of such into relevant data for use in the thesis. For both non-structured interviews and focus groups, open questions are typically used which can make for less reliable findings as they do have the tendency to deviate from the initial question plan, meaning results can vary considerably and replication becomes difficult. For the potential benefit that these options might have brought to the thesis, it was not reasonable enough to continue to pursue this methodological approach.

Doctrinal and interdisciplinary methods

As referenced earlier in this section, the methodology primarily utilised within this thesis is the critical doctrinal approach as it was deemed that this option would yield optimal findings and create the most robust research possible. This type of methodology can be defined as "a critical conceptual analysis of all relevant legislation and case law to reveal a statement of the law relevant to the matter under investigation".[38] This research methodology dominates much of the legal research that we consume.[39] In order to carry out this research, various materials including academic articles, books, government and parliamentary papers, existing statutes and proposals, and expert commentary were critically evaluated in order to identify patterns, weaknesses and gaps in knowledge.

Despite this method being the most common amongst legal researchers, there are those in the field who strongly oppose it, some even going as far as to announce doctrinal research as dead.[40] In his proposition that this methodology is dead, Professor Eric Posner alluded to the idea that this legal research method might remain relevant for legal practice, but not for science.[41] This thesis argues that this is most definitely not the case, whilst the emergence of interdisciplinary research subjects such as AI and law mean that we can be more innovative and wide-ranging in our choices of methods, it does not mean that legal doctrinal research is obsolete. The overall findings of this thesis would not have been possible without the use of this tried and tested method, meaning it is far from 'dead' even in such a novel space.

Saying this, however, there still exists a somewhat strained relationship between traditional 'doctrinalists' and 'interdisciplinarians', who often to struggle to see the overlap and benefit of

[38] T. Hutchinson, 'The Doctrinal Method: Incorporating Interdisciplinary Methods in Reforming the Law' (2015) *Erasmus Law Review* 3 130-138
[39] T. Hutchinson, N. Duncan, 'Defining and Describing What We Do: Doctrinal Legal Research' (2012) *Deakin Law Review* 17(1) 83-120
[40] R. van Gestel, H-W. Micklitz, 'Revitilising doctrinal legal research in Europe: What about methodology?' (2011) *European University Institute Working Papers* https://cadmus.eui.eu/bitstream/handle/1814/16825/LAW_2011_05.pdf?sequence=1&isAllowed=y accessed 22/11/2022
[41] Ibid. Posner stated this at the inaugural conference of the Research Group for Methodology of Lawmaking and Legal Research at Tilburg University in 2008.

using these two methodologies in tandem.[42] The main criticism on behalf of those favouring interdisciplinary research is that those preferring the doctrinal approach are "intellectually rigid, inflexible, formalistic and inward-looking"[43], or in other words old-fashioned. Perhaps this may have been the case at one time, but for the research conducted throughout the course of this thesis, the doctrinal method has been used with a view to findings and proposals being flexible, amenable and broad in their application. This proves that whilst it might appear that doctrinal method can be rather fixed and inflexible, it does not have to be applied in such a manner.

As for interdisciplinarity, it was unavoidable that this method of inquiry would feature within this thesis. Interdisciplinary research is often defined as spanning two or more disciplines or areas of learning[44], the two within this thesis being law and artificial intelligence. Rather than presenting this thesis as an in-depth technical exploration of artificial intelligence as sub-category of the larger computer science subject area, this thesis is written from a legal perspective (as is the expertise of the author). Despite this however, this thesis deals with legal and ethical issues caused as a result of AI design and deployment, and so therefore addressing those issues and presenting technical proposals for minimising these issues was essential. Studying the design, functioning and use of AI within our society was also necessary for the completion of this thesis, in addition to the traditional doctrinal method.

It was also essential to conceptualise how AI specialists and computer scientists are feeling towards and approaching the interaction between AI and the law, as any meaningful regulatory proposals are likely going to impose mandates on designers and programmers in order to dictate how they create these systems. Therefore, bridging the gap between law and AI research was necessary for this thesis to be successful.

Comparative methodology

In recognition of the interdisciplinary, and non-traditional nature of this research, a comparative critical approach was also utilised, specifically in Chapters Three and Four of this thesis. These two chapters feature an analysis of regulatory strategies and approaches to AI in variety of nations (including the UK, US, China, South Africa and Egypt), regions (including the EU and the African region), and the United Nations (as an international body). These jurisdictions were selected due to the diverse variety of AI-related strategies and frameworks being developed within them. This variety of nations, regions and international

[42] Ibid n40
[43] D. Vick, 'Interdisciplinarity and the Discipline of Law' (2004) *Journal of Law and Society* 31(2) 163-*193*
[44] Ibid

bodies each have specific interests they wish to pursue with regards to AI, different priorities and barriers to AI implementation. Therefore, selecting these jurisdictions was done so on the grounds that it presents this thesis with a more rounded look at global preparedness for AI regulation.

It has been argued that the combination of black-letter law analysis, combined with comparative law research is especially ideal for those aiming to suggest reforms to the law, or proposing ways to build upon existing legal principles.[45] Resultingly, a comparative method, paired with the traditional critical doctrinal approach discussed above, has allowed for robust proposals to be made regarding the future of AI regulation in the latter chapters of this thesis.

As per Eberle, in such a globally interlinked world comparative legal studies are necessary in order to identify mutual interests and international commonalities.[46] Especially with regards to regulatory analysis, international harmonisation of approaches is not only desirable but essential in most areas. The comparative approach seeks to understand the role the law plays in other countries, and how we can work together to overcome critical issues in a particular space.[47]

Primarily, the comparative method is most useful for providing critical perspectives that may shape policy development, inform law reform and encourage legal harmonisation.[48] For these reasons, it was logical to utilise the comparative methodology within this thesis. The method features as a central component of this thesis via the critical analysis of jurisdictional approaches contained with Chapters Three and Four, and these chapters go on to form a solid basis for the proposals and recommendations that follow in Chapters Five and Six, and as such are foundational to the robustness of the regulatory proposals made.

As with most methods however, there are some criticisms of this approach and limitations that are necessary to address. The main criticism here is that those using comparative research methods are rather 'obsessed' with finding and creating similarity in the law.[49] Those of this opinion perceive the comparative methodology to be flawed in that it is rather

[45] E. Orucu, 'Methodological Aspects of Comparative Law (2006) *European Journal of Law Reform* 8(1) 29-42 *31*
[46] E. J. Eberle, 'The Methodology of Comparative Law' (2011) *Roger Williams University Law Review* 16(1) 51-72 *60*
[47] Ibid
[48] Ibid
[49] G. Frankenberg, 'Critical Comparisons: Rethinking Comparative Law' (1985) *Harvard International Law Journal* 26(2) *439*

unrealistic, as it is near impossible to ensure exact international harmonisation due to differing cultures and general heterogeneity across states.[50]

However, as per the findings of this thesis, comparative research methods can be used in order to yield realistic outputs. The nature of the research conducted within this thesis which is at the intersection of law and computer science, reinforces the idea that harmonisation and international regulatory collaboration is necessary and possible to an extent. This is because of the integral role that AI continues to play globally; many economies are dependent on its development, and most if not all societies will benefit from it. Therefore, it is essential that any regulatory approaches to this type of technology are harmonised to a degree. It is acknowledged that complete international harmonisation will likely not happen, again due to the heterogeneity of the states looking to regulate AI, but nonetheless it is still necessary to present the benefits of a moderately globally harmonised approach.

Case study method

Another type of methodology utilised in this thesis is the case study methodology, which is usually a classic hallmark of research conducted within social and life sciences.[51] By definition, a case study is an intensive and rigorous investigation of a particular subject or issue with a view to better understanding the subject, and often this investigation will help us to better suggest solutions to an ongoing problem.[52] The case study within this thesis is featured in Chapter Two, and includes an in-depth investigation into the bias and discrimination problem present within AI-based systems. This case study examines four instances of AI-based discrimination and bias, and then considers how well-equipped current UK legal safeguards are in dealing with this issue at large.

This chapter, and in specific the case study method used here, is particularly beneficial to the overall thesis as it justifies the regulatory proposals made throughout this work. By choosing to examine such a negatively impactful and emotive issue as a case study (bias and discrimination within AI can and has affected all manner of people in innumerable ways) means that the proposals that follow are more robust and appropriately supported.

[50] Ibid
[51] R. Heale, A. Twycross, 'What is a case study' (2018) *Evidence Based Nursing* 21(1) 7-8
[52] Ibid

As above, there are various benefits to using a case study methodology. Typically, this method allows us to clearly identify gaps in knowledge and interventions that exist within the examined subject area, and better suggest ways to fill those gaps.[53] Therefore this methodology lends itself well to legal research, in that using a case study to demonstrate the need for legal reform or regulation where there is none already creates strong foundations for arguments to be made.

Selecting the case or subject area to focus on within this method is a crucial task, and one must carefully consider the existing literature, and have some understanding of the issues present within the space before embarking on the case study.[54] There are also several types of case study as set out by Crowe et al, including the intrinsic study, the instrumental study, and the collective study.[55] The intrinsic study is typically were a case is chosen not because it is representative of other cases, but because it is different, meaning it might be of specific interest to researchers in that space.[56] An instrumental study is quite the opposite, this is where a 'typical' case is selected and this might help researchers to develop hypotheses and theories.[57] Finally, the collective study involves choosing a number of cases, and allows comparisons to be made across these cases.[58] This is a useful type of case study approach as choosing two to three cases and comparing their outcomes allows us to make well-rounded assumptions and hypotheses, and makes any resulting proposals and suggestions more robust.

Chapter Two of this thesis utilises the collective study approach; four examples of AI-based bias and discrimination are examined, and comparisons are drawn between them. These cases are then used to provide a basis for the legal analysis that follows in that chapter, and even further as a foundation for the proposals that feature later in the thesis. Choosing the specific cases to examine within a collective case study is integral to the success of the study overall. For this chapter, three initial examples of biased AI were selected to demonstrate the issue; the A-Level results scandal of 2020, the controversial Apple Credit Card, and the discrimination unearthed via the Gender Shades Project. Later in the chapter, the Home Office visa algorithm was also examined as an example of an additional biased AI. These cases were selected for inclusion within this case study as they were lesser-known

[53] S. Crowe, K. Cresswell, A. Robertson, G. Huby, A. Avery, A, Sheikh, 'The case study approach' (2011) *BMC Medical Research Methodology* 11 https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-11-100 accessed 23/11/22
[54] Ibid
[55] Ibid
[56] Ibid
[57] Ibid
[58] Ibid

examples, and most were (at the time of writing) scarcely discussed within the literature. Cases like the COMPAS recidivism tool[59], whilst relevant to the discussion and presenting an excellent example of biased AI, had been covered considerably in literature on the topic of AI-based discrimination already.

As with all research methods, the case study method does have some limitations specifically with regards to reliability and validity.[60] The risk here lies with the collection of too much data. As per Crowe et al, there is a real temptation when conducting a case study to collect as much data as possible, and combined with time constraints, synthesising and utilising that data effectively in the study can be a challenge and might ultimately undermine the studies validity.[61] To avoid this issue, only four cases were selected for inclusion within this study. Whilst these cases clearly demonstrate the bias and discrimination problem present within AI, they do not run the risk of presenting too much data or too many variables which in turn may run the risk of undermining the case study as a whole.

Examining AI through a race, gender and class lens

A recurring discussion throughout this thesis, and one of the most pertinent issues explored within this thesis is that of bias and discrimination in AI-systems. This is discussed in depth in Chapter Two, and is a recurring problem we see time and time again in this field. Therefore, it is important to acknowledge that this thesis examines AI applications in Chapter Two from a variety of lenses including a race lens, a gender lens and a class lens.

The very nature of AI is inherently white, meaning that as it was predominantly and initially developed in a 'white military-industrial-academic complex' it has been built to serve and perpetuate whiteness as an ideology.[62] We have a plethora of evidence that shows us that this is the case, whereby algorithms used within healthcare are effectively deprioritising black patients, and in the criminal justice system they are incorrectly asserting that black offenders are more likely to reoffend.[63] Therefore, it becomes both unavoidable and essential to consider AI from a critical race perspective, for example within Chapter Two this

[59] Angwin, J. Larson, J. Mattu, S. Kirchner, L. 'Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks. ProPublica' (propublica.org, 2016) https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing accessed 23/11/22
[60] L. Krusenvik, 'Using Case Studies as a Scientific Method: Advantages and Disadvantages, Halmstad University' (diva-portal.org, 2016) <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1054643&dswid=189> accessed 23/11/22
[61] Ibid n53
[62] Y. Katz, *Artificial Whiteness: Politics and ideology in artificial intelligence* (Columbia University Press, 2020) Introduction
[63] A. Ahuja, 'Tech luminaries' beliefs need further examination' (ft.com, 2023) https://www.ft.com/content/edc30352-05fb-4fd8-a503-20b50ce014ab accessed 12/05/2023

lens has been used to further the discussion on the impact of inaccurate facial recognition technology.

Not only do we need to evaluate AI from a racial lens but considering AI from a gendered lens has also proved valuable to this thesis. Data science and the development of AI are fields that have always been significantly dominated by men, in particular, white men from the 'global north'.[64] Even in the UK, women make up half of the total population but only 20% of those working within AI and data science.[65] These systems are often reminiscent of their creators and therefore, when we see potentially sexist outcomes resulting from AI (for example the Apple Credit Card as discussed within Chapter Two), it is hardly shocking.

Finally, considering AI from a class lens has also proved to be both important and demonstrable of the large-scale issue in question. Chapter Two considers in detail the A-Level Results scandal in the UK in 2020, an issue caused by a discriminatory algorithm that exacerbated existing social class inequity. Looking at AI through this lens sheds light on the reality that AI may in fact deepen the already existing class divide by placing power in the hands of a small group of 'elite' people who control the algorithms that we are subjected to, and make some jobs usually dominated by members of certain social classes obsolete.[66]

Particularly in Chapter Two, these critical lenses of examination have proved useful tools to assess and analyse the broader impact that AI might have. Whilst acknowledging the discrimination present within the examples discussed in that section, delving deeper into the potential reasoning and causes of such discrimination is fundamental if we wish to find a way to minimise these issues.

Conclusion

This section demonstrates the various research methods and lenses selected for inclusion within this thesis. Rigorous consideration was given to each of the selected methodologies regarding the specific benefits they would bring to the overall work in order to achieve the most robust research possible, and the most innovative original contributions. As per the reasoning provided earlier in this section, the chosen methodologies were deemed most

[64] L. Simpson, 'Looking at AI through a global, gender lens' (medium.com, 2019) https://medium.com/hellobrink-co/looking-at-ai-through-a-global-gender-lens-f12aa92c55a4> accessed 13/03/2023

[65] The Alan Turing Institute, 'Women in Data Science and AI: Project Aims' (turing.ac, 2023) https://www.turing.ac.uk/research/research-projects/women-data-science-and-ai-new#:~:text=Women%20make%20up%20half%20the,online%20and%20physical%20workplace%20cultures. accessed 13/03/2023

[66] I. Sheikhansari, 'The Impact of AI on Social Class and Jobs: A Closer Look' (linkedin.com, 2023) < https://www.linkedin.com/pulse/impact-ai-social-class-jobs-closer-look-iman-sheikhansari/> accessed 25/03/2023

suitable for this thesis over other methods such as interviews and questionnaires, as it was deemed that they would yield the most impactful research necessary for the completion of this thesis.

Chapter One

Introducing AI

1.1 Introduction

Rapid technological development is often accompanied by unforeseen consequences, and this is certainly the case with artificial intelligence (AI). Due to the increased use of AI in all manner of industries, including the use of such by organisations in both the public and private sectors, it is becoming more and more clear that utilising this type of autonomous technology without careful consideration and monitoring can have detrimental impacts.[67] This thesis as a whole considers the need for AI regulation, whilst this chapter in particular closely considers AI use cases that have proved to be a particular cause for concern. This chapter therefore considers the legal and ethical impacts of particular AI use cases, the impacts of which can range anywhere from gender and ethnic-based discrimination[68] to the speculated deconstruction of our democracy,[69] as a basis for better regulation of the technology.

Firstly, it is pertinent to explain what is meant by legal and ethical impacts resulting from AI use and to set the parameters of this work, as these legal and ethical impacts will be referred to throughout this thesis. When referring to legal issues, this work is predominantly concerned with infringements upon legally protected human rights and freedoms caused by AI, for example the right to private life, freedom of assembly and association, and the right to protection from discrimination as enshrined within fundamental legal instruments such as the European Convention on Human Rights,[70] and the Charter of Fundamental Rights.[71] This thesis also considers data protection issues that arise out of AI use, which are typically legislated for by instruments such as the General Data Protection Regulations 2016 for example.[72]

---

[67] D. Helbing, 'Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence and Manipulative Technologies' in D. Helbing (eds), *Towards Digital Enlightenment: Essays on the Dark and Light Sides of the Digital Revolution* (Springer 2019) 47-72

[68] K. Hannah-Moffatt, 'Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates' (2018) *Theoretical Criminology* https://journals.sagepub.com/doi/pdf/10.1177/1362480618763582 accessed 30/10/2019

[69] R. Wilson, 'Cambridge Analytica, Facebook, and Influence Operations: A Case Study and Anticipatory Ethical Analysis' (2019) *European Conference on Cyber Warfare and Security* 587-595

[70] Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950

[71] European Union, Charter of Fundamental Rights of the European Union, 26 October 2012, 2012/C 326/02

[72] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free

When referring to ethical issues, this means issues caused by AI that conflict with key ethical principles. These are commonly agreed upon ethical principles for AI such as transparency, accountability, and non-discrimination, and are can be formally identified via internationally accepted ethical guidelines such as the EU's High-Level Expert Group on Artificial Intelligence (HLEG) Ethics Guidelines for Trustworthy AI,[73] and the SHERPA guidelines on development and use of ethical AI, which expanded upon the work of HLEG.[74] Several standards development organisations have also contributed to the growing list of ethical principles for AI, including the British Standards Institute (BSI) via BS 8611 on the ethical design and application of robots and robotic systems,[75] and the Institute of Electrical and Electronics Engineers (IEEE) via their Global Initiative on Ethics of Autonomous and Intelligent Systems.[76]

This chapter serves as an introduction to AI and several use cases that have proven to be a cause for concern. This chapter also begins to the lay the foundations for considering how we might begin to regulate AI in an effective way, allowing us as a society to reap the copious technological benefits that AI presents us with, in the safest way possible. This Chapter therefore directly addresses the first research question set out in this thesis by investigating some of the most pertinent issues/threats posed by AI at present, whilst establishing solid foundations for the rest of the thesis.

1.1.1 Defining artificial intelligence

To fully appreciate the extensive use of AI within our modern society and the resulting ethical impacts of its use, our grasp of what is understood as AI for the purposes of this thesis should be clarified. If we fail to understand the concept of artificial intelligence from the outset, then it is highly unlikely that we will be capable of effectively regulating it.[77] Therefore,

movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

[73] European Commission, High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (*ec.europa.eu*, 2019) https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html accessed 04/04/2023

[74] SHERPA, 'Guidelines for the Ethical Use of AI and Big Data Systems', and 'Guidelines for the Ethical Development of AI ad Big Data Systems: An Ethics by Design approach' (*project-sherpa.eu*, 2020) https://www.project-sherpa.eu/guidelines/ accessed 04/04/2023

[75] British Standards Institute, 'BS 8611 Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems'(standardsdevelopment.bsigroup.com, 2016) https://standardsdevelopment.bsigroup.com/projects/2022-00279#/section accessed 04/04/2023

[76] IEEE, 'The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems' (standards.ieee.org, 2017) https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_general_principles_v2.pdf accessed 04/04/2023

[77] N. Wirth, 'Hello marketing, what can artificial intelligence help you with? (2018) *International Journal of Market Research* 60(5) 435-439 *436*

a definition of AI is crucial to this text in order to provide context for the legal and ethical discussions that will take place throughout this thesis.

AI by nature is hard to define[78]; our general understanding of AI differs in technicality depending on the field in which the system is considered.[79] Many scholars are of the opinion that AI is significantly hard to define; it is a term used widely by many but truly understood by few.[80] Realistically, the AI that we know today has been a reality only for a limited time, for the past decade or so, due to advancements in data sets and the development of more powerful hardware.[81] Therefore, our understanding of what AI actually is fluid and changes frequently.

Any efforts to define the term seem to differ in technicality based upon the field in which the definition is considered, however, the relatively broad definition provided by Nilsson seems suitable for the purposes of this thesis.[82] He provides that AI is concerned with the 'intelligent behaviour of artefacts', with the term intelligence referring to concepts such as reasoning, communicating, and learning.[83]

The idea that intelligence must be understood in order to fully appreciate artificial intelligence as a whole, is commonly accepted.[84] However, it is held by many that defining intelligence alone is, again, somewhat difficult, considering that this includes trying to draw the line between actual thinking and a purely mechanical process.[85]

This thesis considers AI in its legal and ethical context as opposed to a strictly scientific setting, as a result it is suitable to define artificial intelligence as an umbrella term.[86] Therefore, this thesis adopts the definition of AI as the intelligent behaviour of a device that is artificial (non-human); a device that can carry out tasks, learn from its environment and apply knowledge in a similar way to a human.

---

[78] D. Parnas, 'The real risks of artificial intelligence' (2017) Communications of the ACM 60(10) 27-31 27
[79] N. J. Nilsson, Artificial Intelligence: A New Synthesis (Morgan Kauffman Publishers, San Francisco 1998)
[80] Ibid n65
[81] J. Joseph, U. Turksen, 'Harnessing AI for due diligence in CBI Programmes' (2022) *Journal of Ethics and Legal Technologies* 4(2)
[82] Ibid n79
[83] Ibid n79
[84] Ibid n77
[85] Ibid n77
[86] R. Welch, 'Defining Artificial Intelligence' (2019) Society of Motion Picture &Television Engineers Motion Imaging Journal 128(1) 26-33 26

1.2 Uses of artificial intelligence within our modern society

We are currently residing within what is known and widely referred to as the fourth industrial revolution; or otherwise titled the technological revolution.[87] As such, we live within a world that is becoming increasingly interconnected. Decisions that would have typically been made by humans are being made by automated systems, our online presence is being scrutinised in such a way that our behaviours are being predicted more accurately by an algorithm than by the people we know and interact with on a daily basis.[88] As succinctly, stated by the Organisation for Economic Co-Operation and Development (OECD), artificial intelligence is reshaping our economies and allowing us to make better and more informed decisions.[89] Yet despite this, the growing use of AI in all manner of industries is leading to an increased sense of social anxiety and concern, primarily rooted in the ethics of using these autonomous systems in such impactful ways.[90]

It is interesting to consider just how many 'smart' devices (containing AI) that we come into contact with on a typical day, interactions that we might not usually spare a moment to consider. We may use facial recognition technology to unlock our phones, built in systems within our cars predict our potential destination based upon the time of day and day of the week. We may browse the web or make a purchase online via a website such as Amazon, which will then go on to suggest other products we may be interested in purchasing based on this interaction.

Our daily interactions with intelligent systems and devices are countless. It is not necessary nor reasonable to include an exhaustive list of devices that incorporate AI within them within this chapter in order to demonstrate the sheer number of times each day that we encounter autonomous technology. One commonality that does appear consistent is that despite this constant interaction, the average individual does not fully realise the rate at which artificial intelligence is developing, and as such, we fail to fully appreciate how frequently and the extent to which this type of technology is being implemented into our lives.[91] With this level of technological interaction comes concern regarding the potential safeguards in place to protect us from any undesirable consequences of such boundless AI implementation.

---

[87] K. Schwab, *The Fourth Industrial Revolution* (Crown Publishing 2016) 1
[88] Ibid n31 in 'Preface'
[89] OECD, *Artificial Intelligence in Society* (OECD Publishing, Paris 2019) <https://doi.org/10.1787/eedfee77-en> accessed 02/11/2019
[90] ibid
[91] T. Hauer, 'Society and the Second Age of Machines: Algorithms vs Ethics' (2018) *Society* 55(2) 100-106 *100*

Further, the rapid rate at which AI is developing is unlike anything we have seen before. This growth has been described as 'exponential rather than linear' due to the nature of the technology itself; the development of one autonomous device essentially brings about the development of an even more sophisticated and capable device.[92] The speed with which artificially intelligent technology is evolving also feeds directly into the call for closer consideration of regulatory measures for AI. This is due to the fact that with such exponential growth, the race to create the most cutting edge, intelligent device is on; organisations are more likely to 'cut corners' in relation to safety measures and responsibilities for these systems, in order to become leading names in the AI sphere.[93] There is therefore an accompanying risk that we will begin to see the development of unsafe and flawed AI.

As submitted by Ghandi et al[94], this predilection that all industries should utilise AI to such an extent is a reflection of our intolerance of the unknown; by utilising autonomous intelligent devices it is hoped that we can 'limit daily variability and attempt to optimise productivity'. This is the crux of the issue surrounding the astronomic growth in the use of artificially intelligent systems, particularly by powerful, influential organisations and institutions. These bodies wish to utilise AI in order to increase their own capacities, to broaden their reach to demographics that they may not have engaged with in the past in the hope of increasing productivity and profit, and to put to use the great quantities of information made available to them.

This is inevitable; all 'stakeholders of global society' will eventually see the infinite potential that modern technology such as AI holds for their organisations, and they will develop and deploy AI happily in the absence of meaningful regulation.[95] Therefore, it is vital that we focus not on preventing the use of such technology; this would be counterintuitive and futile, but our focus must be on developing our understanding of the capabilities of artificially intelligent technologies, how we harness that capability, and how we ensure its safe use.

[92] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, 'Artificial intelligence in healthcare: past, present and future' (2017) *Stroke and Vascular Neurology* <https://svn.bmj.com/content/svnbmj/2/4/230.full.pdf> accessed 02/11/2019
[93] S. Cave, S. OhEigeartaigh, 'An AI Race for Strategic Advantage: Rhetoric and Risks' (2018) Association for the Advancement of Artificial Intelligence <http://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_163.pdf> accessed 2 November 2019
[94] S. Ghandi, W. Mosleh, J. Shen, C.M. Chow, 'Automation, machine learning and artificial intelligence in echocardiography: A brave new world' (2018) *Echocardiography* 35(9) 1402-1419 *1402*
[95] Ibid n87

AI in Healthcare

Further to this point, it is useful to consider some of the ways in which stakeholders have been able to use AI in order to enhance performance within their industries and sectors. One sector in which AI has been put to use rather successfully is healthcare. Within this sector, artificially intelligent systems have been used particularly well in relation to patient diagnosis, treatment and prognosis evaluation.[96] The primary purpose for using AI in patient care is somewhat of a reiteration of the submissions made by Ghandi et al[97] and discussed earlier in this section; to lessen the chance of human variance and ultimately error.[98] This is a fundamental aim for any healthcare provider, to decrease the occurrence of error in the care that they provide. However, it is also of great importance that we are able to ensure that any automated decision put into action in this space is accurate and reliable.

Here, AI has the capability to transform the practice of medicine and revolutionise healthcare service provision. To this effect, AI-based systems are being used to assist clinicians in the selection of suitable cancer drugs on a patient-by-patient basis, to identify potential heart disease in patients, and to assist public health officials in identifying infectious disease outbreaks.[99] Yet, recent studies have shown that patient apprehension to this technology pervading medical care is varied and could act as a potential barrier to the continued use and development of AI in this way.[100]

We are therefore seeing societies anxieties towards mass deployment of AI within our every-day lives come to fruition. Questions are asked regarding how reliable and accurate these systems are, and what level of supervision is necessary to ensure their safe use. One major concern here is that any AI used in this setting should be used in a complimentary fashion, that is to say in accompaniment to human clinical practitioners, and not in replacement of these human clinical practitioners.[101] Addressing these concerns is therefore crucial in order to avoid a so-called 'AI Winter' and to encourage public trust in these technologies, the ways in which we might begin to do this will explored further on in this thesis.

---

[96] Ibid n92
[97] Ibid n94
[98] Ibid n92
[99] E. Racine, W. Boehlen, M. Sample, 'Healthcare uses of artificial intelligence: Challenges and opportunities for growth' (2019) *Healthcare Management Forum* 32(5)
[100] J. P. Richardson, C. Smith, S. Curtis, S. Watson, X. Zhu, B. Barry, R. R. Sharp, 'Patient apprehensions about the use of artificial intelligence in healthcare' (2021) *NPJ Digital Medicine* 4 <https://doi.org/10.1038/s41746-021-00509-1>
[101] S. Reddy, J. Fox, M. P. Purohit, 'Artificial Intelligence-enabled healthcare delivery' (2018) *Journal of the Royal Society of Medicine* 112(1) 22-28 *22*

Having explored an example of a positive use case for AI, it is also important to note the rather less effective uses of AI. Ones that have either detrimentally affected members of society or stirred our democratic process considerably.

AI in the criminal justice system

Similar to the use of AI within healthcare, the use of AI within judicial proceedings also has the potential to revolutionise the sector. There are a plethora of ways in which AI may be deployed here, however, for the purpose of this thesis AI use within the criminal justice system will be considered specifically. On the face of it, utilising AI within the criminal justice system has appealing factors such as the possibility to increase efficiency in a number of time-consuming processes and reduce costs, however, occurrence of algorithmic bias in this particular field has the capacity to be considerably damaging and unethical in nature.[102]

As with any other industries or sectors, it was inevitable that AI would eventually find its way into the criminal justice process, and we are now seeing it used by police forces and courts alike in multiple jurisdictions, predominantly to aid decisions such as appropriate sentence length and proportionate bail amounts.[103] It is argued that use of AI-based systems in this capacity are a step towards creating a more efficient, functional and accurate justice system, yet despite this stance, it is being proven to present somewhat disproportionate risks.[104]

Risk assessment algorithms are just one of the initial types of AI being used within the criminal justice system, particularly in the United States of America (US), with extensive accompanying studies being carried out to investigate the accuracy of these algorithms.[105] One system in particular that has caused controversy in relation to its ethical implications is the COMPAS assessment tool (Correctional Offender Management Profiling for Alternative Sanctions).[106] This algorithmic risk assessment system has been in operation since 1998 and in that time has been utilised in order to predict the behaviours of over a million offenders and their likelihood of reoffending based on one hundred and thirty seven specific

[102] For further detail regarding algorithmic bias, please see Chapter 2.
[103] Ibid n19
[104] Ibid n68
[105] Electronic Privacy Information Centre, 'Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools' (epic.org, 2019) https://epic.org/algorithmic-transparency/crim-justice/ accessed 10/11/2019
[106] J. Dressel, H. Farid, 'The accuracy, fairness, and limits of predicting recidivism' (2018) *Science Advances* 4(1) https://advances.sciencemag.org/content/4/1/eaao5580 10/11/2019

features, (including factors such as the offenders' previous criminal record).[107] It does this by producing a risk score for each individual, which indicates their likelihood of reoffending.

Angwin et al conducted a study as to the reliability of this score given by assessing whether or not those predicted as highly likely to reoffend did so within two years following the score being given.[108] A rather disconcerting and unsurprising conclusion became apparent; the score was unreliable, in particular with relation to mistakes made regarding the reoffending rates of black and white defendants. Racial disparities were obvious, white offenders where often scored incorrectly as having a low reoffending probability when in fact it should have been higher, whilst black defendants were wrongly labelled as 'future criminals' at twice the rate as white offenders, which in fact was completely inaccurate.[109]

This study demonstrates the colossal impact that using artificially intelligent systems within the criminal justice process can have, so much so that members of the judiciary in the US have made comment on the use of these risk assessment algorithms. US Attorney General Eric Holder stated in 2014 that despite these intelligent systems being deployed with the 'best of intentions', they should be used with caution so as to prevent bias and inequality being injected into our justice systems.[110] The use of these predictive algorithms poses potential issues surrounding unfairness and inequality for the judiciary, not to mention the impact that a wrongful and inaccurate prediction might have upon an individual's life.

It is to be expected that as we see an increase in the use of AI within our criminal justice systems, as in other sectors, we will also so see a rise in litigation regarding the legality of using these 'close-sourced risk assessment software'; we can see the beginnings of such litigation in the case of *Loomis v Wisconsin*.[111] Interestingly, in this case the risk assessment software used in order to calculate Loomis' six-year prison sentence was questioned in relation to its legality as it was alleged to have violated the defendant's right to due process. This was due to the software being close-sourced, it was not transparent and therefore it violated a defendants' constitutional right to challenge the validity and accuracy of the test applied to him. In addition to this, it was put by the defendant that the software (COMPAS) actually violates due process by taking both gender and race into account. The Supreme Court declined to hear the case.

[107] Ibid
[108] Ibid n59
[109]Ibid
[110] Ibid
[111] 881 N.W.2d 749 (2016)

Loomis v Wisconsin presented an opportunity for the judiciary in the US to better consider the place that algorithms have within the criminal justice system, which they failed to accept upon declining to hear the case. Although the defendant had limited insight into how this algorithm worked, it was still allowed to play a part in the sentencing process of an individual which is extremely concerning.

In *Loomis,* focus was placed on access to the source code of the algorithm, which its creator would not give access to in court, as it was a 'trade secret'.[112] Interestingly though, as argued by Israni, we should be focusing on the data used by the algorithm and the weighting of that data in relation to the decision reached rather than the source code of the algorithm itself; the bias isn't necessarily in the source code of the algorithm but the bias is implicitly contained in the data that powers the software.[113] Therefore, if the algorithm takes into consideration somebodies postal code and this factor has a higher weighting than other factors also considered, there would likely be an implicitly biased result.[114]

It would be counterintuitive and futile to attempt to prevent the use of intelligent systems in areas such as criminal justice; therefore, caution needs to be exercised in order to prevent unfairness and inequality resulting from the use of AI in our justice system, that is not to say however that we should refrain from using this technology completely.[115] All stakeholders must exercise a certain standard of caution and supervision in order to ensure that technology is being used responsibly, however, at present we do not know what that standard looks like.

Data targeting

In addition to the aforementioned AI use cases, it is pertinent to consider the use of data targeting, to which we are all subjected every day. We as a society leave a trail of data almost everywhere we go, whether that be physically or digitally. Through our interactions on social media, the terms we search in various search engines, the items we purchase online, to the videos we spend our spare time watching, our digital footprint grows and grows. As a result, organisations have access to information regarding many of our personal attributes,

[112] ibid
[113] E. Israni, 'Algorithmic Due Process: Mistaken Accountability and Attribution in State v Loomis' (2017) *Harvard Journal of Law and Technology – JOLT Digest* <https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1> accessed 11/11/2019
[114] Ibid
[115] Ibid n59

including and not limited to, what food we like to eat, what bars we like to frequent, what job we have and what jobs our friends and family have, whether or not we are in a relationship, how long we have been in that relationship, if we have children etc.[116] Organisations of every kind therefore have access to a wealth of personal information regarding us, and as such we should consider how they use this data and what they use it for.

It is inevitable that in our increasingly connected society this type of information has and will continue to become easily accessible, as the majority of us offer this information freely via our social media pages for example. Despite this, it does however provide 'data brokers' the opportunity to harvest unprecedented amounts of data relating to our own personal behaviours, data that is then used in order to target us in a variety of ways.[117]

AI is utilised to make sense of this wealth of information, the resulting knowledge imparted by the AI-based system is then used to target and influence a number of our general daily activities.[118] This methodology allows data brokers to form increasingly refined user profiles, profiles which accurately predict our behaviours, our likes and dislikes, the adverts and products we are most likely to interact with, and even our political persuasion.

One of many case studies that demonstrates the sizeable impact that data and behavioural targeting can have is that of the Cambridge Analytica scandal; in which there is evidence of not only a large-scale data breach, but we are presented with the opportunity to understand exactly how some institutions and organisations are using our data in a potentially destructive way.

This scandal received a considerable amount of media coverage at its height, and has shed light upon the occurrence of data targeting and its impacts; one particular clear demonstration of this can be found in the Netflix documentary titled 'The Great Hack', which explores the involvement of Cambridge Analytica in the US Presidential Election campaign

---

[116] C. Dewey, '98 personal data points that Facebook uses to target ads at you' *The Washington Post* (Washingtonpost.com, 2016) https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/ accessed 16/11/2019

[117] A. Gupta, 'The Evolution of Fraud: Ethical Implications in the Age of Large-scale Data Breaches and Widespread Artificial Intelligence Solutions Deployment' (2018) *International Telecommunications Union Journal* 1 <https://www.researchgate.net/profile/Abhishek_Gupta193/publication/323857997_The_Evolution_of_Fraud_Ethical_Implications_in_the_Age_of_Large-Scale_Data_Breaches_and_Widespread_Artificial_Intelligence_Solutions_Deployment/links/5aaffd3f0f7e9b4897c1d066/The-Evolution-of-Fraud-Ethical-Implications-in-the-Age-of-Large-Scale-Data-Breaches-and-Widespread-Artificial-Intelligence-Solutions-Deployment.pdf> accessed 19/11/2019

[118] Ibid

and also the UK's exit from the European Union (the Brexit campaign).[119] In this work, the use of AI in utilising data is examined and described by an ex-employee of the company as being a potential weapon, a privacy risk and threat against democracy.[120]

For context, Cambridge Analytica was a political communications company that made efforts to gather as much data as possible on US citizens in particular, in order to target and eventually influence their 'voting behaviours', all of which was facilitated by lack of adequate protection of personal, private data by both government and social media companies.[121]

In particular, it is interesting to consider Cambridge Analytica's association with Facebook, despite their insistent denial that any kind of connection between the two companies exists.[122] The interaction between these two companies is of interest in particular because of the obvious data breaches that took place during recent years; Facebook allowed Cambridge Analytica to have unlimited access to the personal information of over eighty seven million Facebook users, and then proceeded to use this information in order to target voters.[123]

As early as 2014 Cambridge Analytica had between 'two thousand and five thousand individual data points' on each American adult, data points being pieces of information; therefore, this amounted to potentially two thousand to five thousand pieces of information on over two hundred and forty million people, all categorised into various political categories.[124] This raises a number of ethical questions in relation to the rights that an individual has concerning the data they produce and the way in which that data can and should be used (two clear ethical issues here regard both lack of transparency and the potential for discrimination that arises from such data being collected and used).

When the term 'rights' is used here, it is used to refer to the rights such as those that can be found in a variety of legal instruments such as the Charter of Fundamental Rights of the European Union,[125] the European Convention on Human Rights,[126] the Universal Declaration of Human Rights[127] to name just a few, all of which exist to promote the protection of

---

[119] J. Noujaim, K. Amer, *The Great Hack* (Netflix 2019)

[120] Ibid

[121] B. Kaiser, *Targeted: My Inside Story of Cambridge Analytica and how Trump, Brexit and Facebook Broke Democracy* (Harper Collins, 2019) Prologue

[122] C. Cadwalladr, 'The Cambridge Analytica Files' *The Guardian* (London, 18th March 2018) <http://davelevy.info/Downloads/cabridgeananalyticafiles%20-theguardian_20180318.pdf> accessed 25/11/2019

[123] J. Isaak, M. J. Hanna, 'User Data Privacy: Facebook, Cambridge Analytica, and Private Protection' (2018) *IEEE Computer* 51(8) 56-59 *56*

[124] Ibid

[125] Ibid n71

[126] Ibid n70

[127] Universal Declaration of Human Rights 1948, 217 A (III)

universal basic rights and freedoms. With regards to this example, the rights most at risk might include the right to private and family life, freedom of expression or freedom of association.

As put by Gupta, it is most concerning that this data was obtained without consent and then used in order to target those very same individuals.[128] It is useful to consider the affect, if any, that the data breach had upon Facebook and whether there where any repercussions that may in fact have damaged the company's reputation or their income occurred as a result. Interestingly however, studies have concluded that despite the scandal, the number of individuals using Facebook decreased so minimally that there was almost no effect on Facebook's profits.[129]

This demonstrates that organisations like Facebook (or Meta) will likely not be discouraged from committing data breaches based purely on the possibility of a decrease in profits.[130] Despite the fact that the public had knowledge of this breach, the actual reduction in users was so incredibly nominal, it had very little impact at all. As a result, this allows us to reach the conclusion that regulation that is more stringent must exist in order to prevent these incidents becoming a regular occurrence.

One question arising from this incident is to what extent might an event such as this contribute to an increasingly splintered society, and what affect (if any) might this have upon our democracy.[131] The answer is uncertain, but there is definitely evidence to suggest this form of targeting is detrimental in numerous ways. It is in fact becoming more common for political campaigns to look more like 'traditional consumer marketing' than actual politically charged campaigns.[132]

This is typically due to the collection of consumer data, which is then processed and regurgitated back to the public, resulting in the occurrence of fake news stories and incorrect information online, in the hope that this will in some way influence voting choices.[133] This was particularly the case in the 2016 presidential election campaign, in which US voters likely to vote for Democrat candidate Hillary Clinton, were more likely to see messages on

---

[128] Ibid n117
[129] J. L. Edwards, 'An Examination of Consumers' Social Media Trust in the Wake of the Facebook and Cambridge Analytica Scandal' (cahe.kzoo.edu, 2019) <https://cache.kzoo.edu/handle/10920/36677> accessed 25/11/2019
[130] See Chapter Six for further detail regarding the effectiveness of punitive sanctions in encouraging regulatory compliance.
[131] S. Morgan, 'Fake news, disinformation, manipulation, and online tactics to undermine democracy' (2018) *Journal of Cyber Policy* 3(1) 39-43 *39*
[132] Ibid
[133] Ibid

their social media pages stating that the date for the election had in fact changed for example.[134]

As a result, it is clear to see the detrimental impact that data targeting can have on individuals with regards to mass amounts of personal data being harvested and used without consent, to the damage that data targeting has the capacity to cause to our democratic processes. Once again, the general response to this pertinent issue is that vigilance is key; we must ensure the safe development and use of AI.[135]

1.3 Cause for regulation

Examples such as the COMPAS recidivism tool have served a purpose in encouraging us to consider what meaningful and effective AI regulation might look like. This means considering whether a self-regulatory approach (relying upon on the stakeholders themselves to ensure standards are adhered to) or a more centralised governance is approach is necessary. Both approaches have benefits and drawbacks,[136] for example, industry self-regulation often has its challenges and typically works best when used to compliment centralised government policies.[137]

A self-regulatory approach to regulating AI would rely on a number of factors including most importantly the willingness of stakeholders to adhere to a set of agreed upon commitments, and to be bound by agreed upon consequences for not adhering to those commitments.[138] This is conceivably the most onerous task in ensuring successful self-regulation, guaranteeing the strength of regulatory measures, finding incentive to ensure adherence to the instruments and also establishing accountability for lack of adherence.[139]
Despite this though, making use of self-regulation were possible does have its advantages; it is more cost effective than government-lead regulation and is arguably more flexible than centralised governance.[140] Therefore, it would appear as though self-regulation is a viable option for governing the use of AI, and in particular preventing the occurrence of algorithmic

[134] Ibid n121
[135] R. M Glassman, 'Will Artificial Intelligence (AI) Make Democracy Irrelevant?' in The Future of Democracy (Springer, 2019) 189-198
[136] See Chapter Five and Six for further analysis on this point.
[137] OECD Directorate for Science, Technology and Innovation Committee on Consumer Policy, 'Industry Self-Regulation: Role and Use in Supporting Consumer Interests' (oecd.org, 2015) <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/CP(2014)4/FINAL&docLanguage=En> accessed 14/11/2019
[138] Ibid
[139] Ibid
[140] Ibid

bias, or at least lessening the impacts of such (this is investigated further in Chapters Five and Six).

As with algorithmic bias, the inherent risks presented by data targeting are also worth consideration. Although data targeting may not result in the same outcomes as algorithmic bias, the consequences are still concerning; the biggest risks are arguably increased large-scale data breaches and the potential detrimental widespread impacts that these might have upon society.

Some argue that in fact the answer to preventing data breaches is AI itself.[141] AI presents several interesting opportunities for creating more secure systems that prevent cyber-attacks and as a result, breaches of personal data.[142] This is a rather interesting consideration; however, as discussed by Jiang et al, the creation of one sophisticated and intelligent device typically brings around the creation of another even more sophisticated and able AI.[143] Therefore, it is likely that with the development of secure software in order to prevent cyber-attacks and breaches, another even more intelligent system would be evolve with the capability to circumnavigate the original software's once 'secure' security features. It is suggested that to tackle the legal implications of data targeting (such as large scale data breaches and illegal data use) data protection laws such as GDPR etc. must better protect seemingly 'anonymised' data (this is investigated in further detail in Chapter Two).[144] This would in essence mean making provisions for data protection law to better encompass the uses of data, and thus help to prevent its unethical use.

There are therefore several common legal and ethical issues we see arising from AI use, as demonstrated within the aforementioned examples. Legally, we are seeing issues regarding data protection (including unethical data collection and use) as well as human rights infringements (including violations of rights such as that of a fair trial). Ethically, there are three common principles for AI that are in need of attention, these are accountability, transparency, and non-discrimination. These ethical principles come into play in almost all examples of AI 'gone wrong', which therefore means that embedding accountability,

[141] Egress Software Technologies, 'The Future of AI in Data Protection: What do the Experts Say?' (egress.com, 2018) <https://www.egress.com/artificial-intelligence-for-data-protection> accessed 01/12/2019
[142] ibid
[143] Ibid n92
[144] For further information on suggestions for amending current legal safeguards, please see Chapter 2.

transparency and non-discrimination into any AI regulation and governance measures is not only desirable, but essential.

1.4 Conclusions

Effectively regulating the various ways in which we use AI will not be simple, and as such this thesis will build upon the initial points raised within this Chapter to provide reasonable suggestions and proposals for AI regulation. The use cases discussed within this Chapter present the intrinsic detrimental impacts that AI can have upon society; whether that be via racial discrimination as explored in relation to the COMPAS risk assessment tool, via large-scale data breaches, or via potential threats against our democratic processes.

The incorporation of AI into daily life is inevitable and should be encouraged were necessary. Therefore, we are presented with the task of ensuring that we are able to reap the benefits of AI whilst protecting against its unreliable nature. This chapter therefore establishes some of the more notable and recent issues/threats posed by AI in answer to the first research question set out within this thesis, and provides a good foundation for the detailed analysis that follows in Chapter Two, in which one specific threat is focused on; bias and discrimination within AI.

Chapter Two

A Case Study: Risk of discrimination in AI systems

2.1 Introduction

Thanks to the growing availability of data, it has become an invaluable commodity in today's society. The ability to access and use the information made available via large data sets is a skill that is now critical to the success of an organisation.[145] Therefore, it has become common practice to use artificially intelligent systems to process this data; as a result, we are able to identify indicators and predictors present within the data that can be used within automated decision-making processes.

The use of such intelligent technology has meant that interpreting data and using the outputs to make informed decisions can now be done autonomously, typically by a decision-making algorithm as opposed to a human. By algorithm, we are referring to a set of rules or instructions that are to be followed, typically by a computer, in order to complete a problem-solving task.[146] This technological advancement has benefits that are far-reaching; for example, time-consuming and time-sensitive tasks can be undertaken within a fraction of the time it would have previously taken to complete.

The popularity of using algorithmic decision-making systems is neither a new nor an unfamiliar concept, and we are subject to the outcomes of automated decision-making processes every single day. These systems are used to assess credit card applications, sift CVs within recruitment processes, to aid in judicial decision-making, and in medical settings to confirm diagnoses, to name just a few instances.

As with any significant technical development, there are always ramifications, and in this instance both legal and ethical questions arise. Use of AI, particularly within automated decision-making, has a recognisable history of discrimination and bias,[147] and those who are subjected to these automated decisions are at risk of falling victims to a largely unfair and inequitable process.

This chapter analyses the bias and discrimination problem present within AI systems as a case study, which was introduced briefly in Chapter One. This Chapter addresses and

---

[145] Janssen, M. van der Voort, H. Wahyudi, A., 'Factors Influencing Big Data Decision-Making Quality' (2017) *Journal of Business Research* 70, 338–345 <http://www.sciencedirect.com/science/article/pii/S0148296316304945> 23/11/2022

[146] Puntambekar, A.A., *Design & Analysis of Algorithms* (2010, Pune: Technical Publication)

[147] Whittaker, M. Alper, M. Bennett, C.L. Hendren, S. Kaziunas, L. Mills, M. Morris, M.R. Rankin, J. Rogers, E. Salas, M. West, S.M., 'Disability, Bias, and AI' (ainowinstitute.org, 2019) <https://ainowinstitute.org/disabilitybiasai-2019.pdf?fbclid=IwAR31dX3o_nkVf-cirQ9P-yJqRRkT1vcKU3MgcEAeWVwUgA0Ue1c-60Zd9OE> accessed 23/11/2022

answers the first and second research questions set out in this thesis, these are as follows: what are the most pertinent issues and threats posed by AI (a specific focus is given to the discrimination and transparency here), and how well equipped are current legal instruments in dealing with the issues posed by AI?

The first part of this chapter considers what bias within AI systems actually looks like, various examples of such bias, and why this bias leads to discrimination. When considering these examples here, AI is examined through a variety of critical lenses in order to delve deeper into the origin of the different kinds of bias we see present in these algorithms. Here the question of transparency is also briefly evaluated; in particular, how a lack of transparency often exacerbates the issue of algorithmic discrimination. The second part of this chapter features the central focus of this case study: an evaluation of current legal safeguards and their effectiveness in tackling AI-based discrimination.

There are a number of highly important issues that we face due to the increased use of AI-based systems, and the risk of bias and discrimination within these systems is arguably the most concerning of all. Therefore, it is necessary to examine this issue in ample detail to give further reasoning and endorsement for the proposals for regulation that come later in this thesis.

Further to this, the analysis of these safeguards is essential, as these legal measures are the first line of defence in dealing with algorithmic bias and discrimination. This chapter focuses part of its analysis on the legal protections awarded by the General Data Protection Regulations (GDPR), in particular Article 22 of the GDPR, and also the effectiveness of current anti-discrimination law relevant in the UK and Europe, including The European Convention of Human Rights (ECHR) and the Equality Act 2010.

The chapter then moves on to consider these legal safeguards in line with the action brought by the Joint Council for the Welfare of Immigrants and Foxglove against the Home Office. This case was one of the first legal challenges to the use of algorithms in the UK; therefore, this work evaluates the effectiveness of current legal safeguards such as GDPR and relevant anti-discrimination laws and how they were applied in this case. Following this, the chapter ends in summary by considering a number of the ways in which we can deal with the lack of transparency and resulting bias and discrimination problems present within AI, as highlighted throughout the chapter.

2.2 What does bias within AI systems look like?

Thanks to the immense capability shown by intelligent systems, it is often presumed that when AI is deployed within a decision-making context, human input is no longer required or

present due to the devices' ability to "change, adapt and grow".[148] This is partly because we expect that decisions that are made by computers will be based purely on fact and nothing else.[149] Yet the opposite is actually true; bias that is present within decision-making algorithms is often there due to human bias existing within the data already, and the algorithm then continues to bolster this existing bias.

Therefore, the assumption that AI-powered automated decision-making "takes the place of human discretion" is not entirely true.[150] Thus, in order to begin to rid these systems of bias, it is important to consider the bias present within algorithmic decision-making processes in further detail and to examine why this bias is present and why its presence ultimately leads to discrimination.

2.3 What is bias?

As with the term artificial intelligence, the term 'bias' also has a number of meanings depending upon the context in which it is considered. The term bias has the capacity to be applied in a neutral context or alternatively with a "significant moral meaning".[151] This is because in its most simple form, 'bias' means an inclination to choose one thing over another, and, given this, the term can be applied in a neutral context. An example of this provided by Friedman and Nissenbaum is that of a person purchasing ripe fruit over damaged fruit.[152] The person is 'biased' because they chose the ripe fruit over the damaged fruit.

However, compare this to a person refusing to hire somebody based upon their ethnicity; the person is still 'biased', but here there is a significant moral meaning behind the term. Therefore, for the purpose of this work, it is worthwhile considering the term bias within a moral and ethical context. As such when we consider bias, we are considering it in relation to its AI counterpart, typically known as algorithmic bias (which results in unfair and often discriminatory outputs).

Discrimination resulting from the use of automated decision-making is typically the consequence of biases already embedded in the data used to train an algorithm. As put by

---

[148] House of Lords Library, 'Predictive and Decision-Making Algorithms in Public Policy' (lordslibrary.parliment.uk, 2020) p.1 <https://lordslibrary.parliament.uk/research-briefings/lln-2020-0045/> accessed 23/11/2022

[149] Tolan, S. 'Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges' (2019) *Cornell Computer Science* <https://arxiv.org/abs/1901.04730> accessed 23/11/2022

[150] Doleac, J. L. Stevenson. M. T. 'Algorithmic Risk Assessment in the Hands of Humans' (2019) p. 1 <https://ssrn.com/abstract=3489440> accessed 23/11/2022

[151] Friedman, B. Nissenbaum, H, 'Bias in Computer Systems' (1996) *ACM Transactions on Information Systems.* 4(13), 330–347, 332

[152] Ibid

Shrestha and Yang, this means that historic prejudices and stereotypes are perpetuated by the new, automated system.[153] Unfortunately, the groups of people most affected by this are those from minority backgrounds, who are discriminated against based upon characteristics such as gender, race, and even socioeconomic factors such as place of residence and where they attended school. This makes for unfair decision-making processes that are making life-altering judgements based upon factors and indicators that are unnecessary and are to some extent unlawful to consider. As a result, there are many instances in which individuals have been discriminated against based upon these various protected characteristics; a few recent examples of this will be examined in the next part of this chapter.

The scholarship on the topic of algorithmic bias is abundant, and it is widely acknowledged that there is a bias and discrimination problem when it comes to algorithmic decision-making. There appears to be what some describe as a growing suspicion on behalf of both the public and those within academia as to the fairness of algorithms used to make important decisions.[154] To a certain extent, it can be argued that this mistrust is encouraging as it means that there is growing awareness and possibly less tolerance of AI-based systems that are less than transparent in the way they operate.

This inherent problem acts as an obstacle to real progress in using algorithmic decision-making on an even wider scale. This is because deep-seated and well-established prejudices are merely echoed back at us via this modern technology, leaving these biases to remain.[155] Therefore, it is essential that policymakers and regulators are aware of the risk posed by the bias that is present within automated systems, and the cause and scale of such bias if we are to have a chance at minimising the issue.

2.4 Examples of discrimination caused by algorithmic decision-making

There is a plethora of cases of algorithmic decision-making gone wrong, and these span across a wide range of sectors. Examples can be found within education as seen at St Georges Medical School,[156] within recruitment as found with Amazon's discriminatory CV

---

[153] Shrestha, Y.R. Yang, Y, 'Fairness in Algorithmic Decision-Making: Applications in Multi-Winner Voting Machine Learning, and Recommender Systems' (2019) *Algorithms,* 12(9), 199–227

[154] Ibid n117

[155] Rovatsos, M. Mittelstadt, B. Koene, A., 'Landscape Summary: Bias in Algorithmic Decision-Making' (GOV.uk, 2019) *Centre for Data Ethics and Innovation* <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf> accessed 23/11/2022

[156] Lowry, S. Macpherson, G., 'A Blot on the Profession' (europepmc.org, 1988) *British Medical Journal.* <http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2545288&blobtype=pdf> accessed 23/11/2022

sifting tool,[157] and within the criminal justice system as evident via the COMPAS recidivism tool.[158] These are well-known examples of algorithms discriminating against individuals and most of which have been discussed repeatedly throughout the literature already. This chapter, however, discusses three less familiar occurrences of AI-based discrimination; the calculation of A-level results in the UK during 2020, applications for the Apple credit card, and facial recognition software used by Microsoft, IBM, and Face ++. The impacts of these case studies are broad in scope, ranging anywhere from being given a lower credit limit, to missing out on the opportunity to attend university. It is therefore clear to see just how far-reaching and extensive the impact of using biased data to power an automated decision-making process can be.

2.4.1 UK A-level results 2020

In a year of disarray, the release of A-level results in the UK during 2020 caused controversy. Due to the Covid-19 pandemic, students were unable to sit the exams that would usually determine their A-level results. Instead, an algorithm was used in order to assist in the calculation of student grades. This model took into account a number of factors, including historical data regarding past attainment of students at the same school, attainment of this year's students, and the results of past students at the same school in the same subjects.[159] These data were used alongside predicted grades that teachers were asked to formulate for their students and the order in which a teacher ranked their students from highest achievers to lowest achievers.[160] The results of this algorithmic decision-making process led to 39.1% of students being downgraded from their predicted grade in England.[161] The knock-on effect of this is that many students lost their conditional offers to attend university after being considerably downgraded.

Bias was demonstrated after it was discovered that private school students seemed to have fared much better than their comprehensive school counterparts; with 48.6% of private school students scoring A grades and above, whilst only 21.8% of comprehensive students scored the same.[162] This was because more weight was given to teacher-predicted grades

[157] Maedche, A. Legner, C. Benlian, A. Berger, B. Gimpel, H. Hess, T. Hinz, O. Morana, S. Söllner, M., 'AI-based Digital Assistants: Opportunities, Threats, and Research Perspectives' (2019) *Business and Information Systems Engineering.* 61(4), 535–544
[158] Ibid n59
[159] The Telegraph, 'A-Level and GCSE Results Update: How Are 2020 Grades Being Calculated without Exams?' (telegraph.co.uk, 2020) <https://www.telegraph.co.uk/education-and-careers/2020/08/18/a-level-gcse-results-grades/> accessed 23/11/2022
[160] Ibid
[161] The Guardian, 'A-Level Results Day 2020 Live' (theguardian, 2020) <https://www.theguardian.com/education/live/2020/aug/13/a-level-results-day-2020-live-students-teachers-government-ucas-mock-exams-triple-lock-nick-gibb> accessed 23/11/2022
[162] Ibid

when students were in smaller classes.[163] Consequently, private schools benefited from the decision made by Ofqual due to their small cohort sizes, whilst state schools with larger cohorts were disadvantaged. As a result, it was made evident that students attending school in less affluent areas were being wrongfully discriminated against. Further transparency regarding the weighting of predicted grades with regard to class size in the final decision-making process would have perhaps provided further necessary insight at the outset.

In a country where access to higher education is already a contentious topic for debate, this outcome provides clear evidence that when using a predictive decision-making, attention must be paid to the data powering the algorithm, and the potential impact of using such. Teachers' assessments and predictions made for students based upon actual classwork and previous observed performance seemed to have less weighting compared to the judgements made by the algorithm using school-wide historic data, which appears to be unjust and unfair.

Following this incident, the government responded by acknowledging the error made. As a result, teacher-predicted grades would again take precedent over the controversial downgraded algorithm-predicted results. This response is an interesting one; the algorithm was specifically designed with its primary goal being to prevent grade inflation.[164]

Therefore, the motive behind utilising such an algorithm is sound and reasonable; however, the issue is that the algorithm was designed specifically to meet the goal of producing exam results that were free of grade inflation. This meant that the algorithm paid less consideration to other factors and resulted in the bias exhibited here.[165] Despite this U-turn, the fact remains that in this particular case an algorithmic decision-making tool was relied upon with seemingly minimal testing and scrutiny.

Considering AI through a social class lens is therefore illuminating, as it helps us to both recognise and evaluate the potential causes for the level of bias observed in this example. As discussed in the methodology section earlier in this thesis, there is a huge potential for AI to deepen the already existing class divide that exists in countries like the UK, by creating an uneven distribution of power in terms of AI development, and by making certain types of employment obsolete.[166]

---

[163] Goodier, M. 'Top A-Level Grades Soar at Private Schools as Sixth Form Colleges Lose Out' *The New Statesman* (newstatesman.com, 2020) <https://www.newstatesman.com/politics/education/2020/08/top-level-grades-soar-private-schools-sixth-form-colleges-lose-out> accessed 23/11/2022

[164] Jones, E. Safak, C. 'Can Algorithms Ever Make the Grade?' (adalovelaceinstitute.org, 2020) <https://www.adalovelaceinstitute.org/can-algorithms-ever-make-the-grade/> accessed 23/11/2022

[165] Ibid

[166] Ibid n78

It is typical that those with both access to the most innovative technologies and a significant amount of money are able to leverage such technology to improve their position by making themselves richer.[167] One example of this offered by Lutz is the Winkelvoss brothers, already from very privileged backgrounds, they were able to further their position by making considerable investments in cryptocurrency and therefore becoming the first crypto billionaires.[168] Add to this to the fact that within our societies we suffer from various digital inequalities, with some societies around the world faring worse than others, for example those within the global south compared to those within the global north and western hemisphere. These inequalities come in several different forms, such as; access to internet, access to technology such as phones, laptops and computers, and lack of skills in relation to how to use both the internet and the technology itself.[169]

Similarly, a high number of what we might term entry level jobs into professional industries, alongside more manual labour jobs typically carried out by humans, will increasingly become automated over time. This therefore means that 'soft skills' such as communication and leadership will become more sought after. As discussed by Vincent in his article on this topic, these are skills often prioritised by independent schools and universities, meaning that going to an independent school and/or attending university will typically allow one to increase this skillset.[170]

Therefore, if those creating decision-making algorithms like the one discussed here are often already from privileged backgrounds, and those algorithms have been shown to favour a certain part of society, it is evident that AI has the capacity to exacerbate and deepen the already existing class-divide. This becomes all the more insidious when we consider the algorithm used to calculate A-level results, which played a part in whether or not students could attend university; as established, this might determine the type of skillset a person has the ability to develop, and thus their future employability in an increasingly automated world. Therefore, the consideration of AI through a social class lens is both intriguing and worrying and warrants further research in this space.

2.4.2 Apple credit card

Gender-based discrimination is hardly a new phenomenon when it comes to risks related to using AI-based systems, and a recent reminder of this is the reported sexist algorithm used

[167] C. Lutz, 'Digital inequalities in the age of artificial intelligence and big data' (2019) *Human Behaviour and Emerging Technology,* 1 141-148, *144*
[168] Ibid
[169] Ibid
[170] J. Vincent, 'Robots and AI are going to make social inequality even worse, says new report' (theverge.com, 2017) < https://www.theverge.com/2017/7/13/15963710/robots-ai-inequality-social-mobility-study> accessed 20/03/2023

to calculate credit limits for the new Apple credit card.[171] This concern was raised by both David Heinemeier, a prominent tech developer, and Steve Wozniak, one of Apple's co-founders.[172] In this instance it was found that despite having a better credit rating, Heinemeier's partner was denied a request to increase her credit limit, but her husband's credit limit was able to remain 20 times higher. In response to this, the New York State Department of Financial Services stated that they would investigate the workings of the card application system in order to determine whether or not the algorithm violated any financial regulations.[173]

Issuing bank Goldman Sachs released a statement affirming that the algorithm did not actually recognise gender as a factor in its decision-making process.[174] This was confirmed when the case was litigated in March 2021 in which the New York State Department of Financial Services established that the issuing bank did not actually discriminate against applicants based upon their gender.[175] They stated that this was due to the algorithms used not considering 'prohibited characteristics'.

Algorithmic discrimination can still occur however even if the algorithm in question does not recognise a particular variable as a factor for consideration, for example, gender. Despite an algorithm not recognising gender as an indicator, other available information about an individual such as purchases they've made or the job title that they hold can still be used to infer that a person is potentially of a certain gender and thus cause indirect discrimination.[176]

Indirect discrimination is typically where a policy or rule is applied in a uniform way for everyone, but disproportionately affects a group of people that share a protected characteristic and puts that group at a disadvantage.[177] This is intrinsically different from

---

[171] BBC, 'Apple's "Sexist" Credit Card Investigated by US Regulator' (BBC.com, 2019) <https://www.bbc.co.uk/news/business-50365609> accessed 23/11/2022
[172] The New York Times, 'Apple Card Investigated after Gender Discrimination Complaints' (nytimes.com, 2019) <https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html> accessed 23/11/2022
[173] Ibid
[174] Knight, W.,The Apple Card Didn't 'See' Gender—And That's the Problem (2019) *Wired* <https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/> accessed 23/11/2022
[175] New York State Department of Financial Services, 'Report on Apple Card Investigation' (dfs.ny.gov, 2021) <https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf> accessed 20/03/2023
[176] Berg, T. Burg, V. Gombović, A., 'On the Rise of the FinTechs — Credit Scoring Using Digital Footprints' (fdic.gov, 2018) Federal Deposit Insurance Corporation <https://www.fdic.gov/bank/analytical/cfr/2018/wp2018/cfr-wp2018-04.pdf> accessed 23/11/2022
[177] Equality and Human Rights Commission, 'What is direct and indirect discrimination?' (equalityhumanrights.com, 2019) < https://www.equalityhumanrights.com/en/advice-and-guidance/what-direct-and-indirect-discrimination> accessed 23/11/2022

direct discrimination in which a person or group is treated worse than others due to having a protected characteristic.[178]

Indirect discrimination, or structural discrimination, is just as impactful as direct discrimination as it still puts an individual at a disadvantage even though a neutral practice or provision might be in place.[179] A good example of this, as provided by Žliobaitė,[180] is that of a person being required to show a driver's license as a form of ID; whilst this might be a neutral requirement, this indirectly discriminates against those with visual impairments who cannot obtain a driver's license. The same principle applies to indirect discrimination based upon gender. The inclusion of so-called proxy information within a data set means that information such as a person's height, which can correlate with gender, might mean that they are still discriminated against based upon their sex.[181]

Further evidence of indirect discrimination based upon gender can be observed within an experiment conducted by Nikhil Sonnad. In this particular experiment, Sonnad used Google translate to translate words from Turkish into English with the Turkish text being in a neutral third-person form.[182] Google Translate assigned the words with a gender, such as 'hard-working' with 'he' and 'lazy' with 'she'. Therefore, the algorithm revealed its apparent gender bias.[183] Whilst the gendering of words is common within many languages, the introduction of AI in this context has actually demonstrated the pre-existing bias within these algorithms.

It is arguably just as harmful to 'turn a blind eye' to an important identifier such as gender, as it is to ignore it. If an algorithm is taught to recognise a factor such as gender, and this algorithm creates sexist outputs, it will be easier to identify bias within the system and to stop this from occurring again in the future, than if the system didn't recognise gender at all.

Further to this discussion on discrimination by proxy, it is of benefit to consider AI from a gendered lens in this respect. As referenced previously in the methodology section of this thesis, the field of computer and data science is dominated by white men.[184] The number of women working within this field is increasing but it there is still a considerable

---

[178] Ibid
[179] Žliobaitė, I. 'A Survey on Measuring Indirect Discrimination in Machine Learning' (arxiv.org, 2015) *Cornell Computer Science* <https://arxiv.org/abs/1511.00148> accessed 23/11/2022
[180] Ibid
[181] European Union Agency for Fundamental Rights, '#BigData: Discrimination in Data-Supported Decision Making' (fra.europa.eu, 2018) <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf> accessed 23/11/2022
[182] Sonnad, N., 'Google Translate's Gender Bias Pairs "He" with "Hardworking" and "She" with Lazy, and Other Examples' (qz.com, 2017) *Quartz* <https://qz.com/1141122/google-translates-gender-bias-pairs-he-with-hardworking-and-she-with-lazy-and-other-examples/> accessed 23/11/2022
[183] Wellner, G. Rothman, T., 'Feminist AI: Can We Expect Our AI Systems to Become Feminist?' (2020) *Philosophy & Technology.* 33, 191–205, 192
[184] Ibid n69

disproportionate representation of women in this space. Afterall, AI is what we make it; meaning that those creating algorithms ultimately shape what the algorithm looks like, as does the data we use to train it. It is therefore relatively clear why we might see that women are disproportionately discriminated against by AI when compared to men.

If we consider historically the role of women in society, women either stayed at home as opposed to taking on work, or more recently we see a lack of representation in women in senior roles within organisations, alongside women still receiving unequal pay.[185] Therefore, despite obvious progression in terms of gender equality in recent decades, there still exists some notion of women as 'lesser than', and if this notion is either consciously or subconsciously perpetuated by a programmer, we may find a biased AI system as the result. Similarly, especially in the case of AI used for recruitment purposes, if the data shows that historically successful applicants for the role of a software engineer at an organisation have been white men, then the AI will be trained to seek out these candidates over women and other groups.

Therefore, for us to truly combat this issue of gender-based discrimination we must continue to look at AI through a gendered lens. It is evident that to minimise this specific issue we need to increase diversity in the AI/data science workforce, and to consider how to reverse the historical preconceptions a considerable number of people still possess with regards to women.

2.4.3 Facial recognition—Microsoft and IBM

Facial analysis software aims to recognise an individual based upon their appearance. This type of software can make predictions regarding a person's gender or race, but the possibilities are seemingly endless. As discussed by Buolamwini and Gebru,[186] variations of this software have claimed to be able to identify emotions and even an individuals' sexuality based on images of the subject. Research carried out by the Gender Shades project identified that Microsoft, IBM, and Face ++, all of whom offer 'gender classification products', had difficulties in identifying subjects accurately.[187]

This project grouped subjects by gender and skin type. As a result, "bias in this context is defined as having practical differences in gender classification error rates between

[185] M.C. Jackson, 'Artificial Intelligence and Algorithmic Bias: The Issues With Technology Reflecting History & Humans' (2021) *Journal of Business & Technology Law* 16(2) 299-316, 309
[186] Buolamwini, J. Gebru, T., 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' (2018) *Proceedings of Machine Learning Research*. 81, 1–15
[187] Gender Shades 'Overview' (gendershades.org, 2018) <http://gendershades.org/overview.html> accessed 23/11/2022

groups".[188] In this particular instance, it was found that Microsoft had difficulties in correctly identifying the gender of darker-skinned subjects, IBM struggled with identifying darker-skinned females, and Face ++ frequently 'misgendered' female subjects.[189] Interestingly, both IBM and Microsoft responded to these findings by stating that their gender classification software would undergo investigation, whereas Face ++ did not respond.[190] The lack of transparency provided in instances such as this, particularly with regard to Face ++, is yet another reason for the growing distrust of artificially intelligent systems.

This project highlights the lack of neutrality in AI systems that we often falsely assume exists.[191] If we are to use facial analysis software far and wide, it needs to be accurate and regulated so as not to fall into the hands of those who will abuse it. When past biases and preconceptions of those who create these AI-based systems are present, the risk of discrimination based on protected characteristics is high.

For this reason, it has been particularly important to look at AI via a critical race lens. As put by Katz in his seminal work on this topic, AI is a technology of whiteness.[192] From its birth in the collaboration between the military and academia, AI was both named and heralded by predominantly white men in the global north. Katz goes further in asserting that "Whiteness has been shaped by the need to accumulate land, maintain the supply of unfree labour, and in settler-colonialist societies, erase indigenous peoples".[193] This is a powerful and accurate depiction of whiteness as a concept, and also explains somewhat the discrimination we see within AI systems.

As time has gone on, we see that whiteness as a tool for control has changed; for example, we mightn't see belligerent white supremacy as we once did throughout history, but the control and power that whiteness as a concept has is now manifested in different ways for example via our criminal justice systems, our government housing schemes etc.[194] Therefore AI simply mirrors and perpetuates this manifestation of whiteness, making it all the more clear to see why these systems create disproportionate and discriminatory outcomes against those who are not white.

[188] Ibid
[189] Ibid
[190] Ibid
[191] Ibid
[192] Ibid n67
[193] Ibid n67
[194] Ibid n67

As demonstrated earlier in this section, if historical prejudices remain and disproportionality exists within the data we use to train these systems, then we are left with antiquated AI.[195] Team this with the reality that AI is fundamentally used to further the agenda of predominantly white capitalist societies and we can clearly see the cause of such discrimination and bias. This therefore leads us to consider how exactly we might hope to combat this issue, perhaps by acknowledging the clear link between AI and whiteness might be a start to consider how truly trustworthy AI is in its current state. More on this discussion will feature in Chapters Three and Four of this thesis.

2.5 Lack of transparency

It would therefore appear that many of the issues relating to discrimination within AI-based systems are the result of, or are exacerbated by, a lack of transparency, i.e. lack of information regarding the data and how the system is using the data. As we become increasingly subject to the automated decision-making process, it is expected that there will be further questioning regarding the use of algorithms and the processes undertaken to reach such decisions. To answer these questions, we need more transparency in AI-based systems.

In fact, during the summer of 2020, two UK drivers commenced legal action against their employer Uber, requesting access to their personal data and requesting transparency regarding the use of automated decision-making within their employment.[196] They alleged that these automated decisions have an impact upon the jobs that drivers are allocated and the pay that they receive.[197] This case is clear evidence that transparency is a growing concern with regard to automated decision-making, and to ensure trust in the process we need to ensure transparency first.

Finck discusses that as automated decision-making begins to take the place of human decision-making, administrative law principles, such as transparency, are likely to be challenged.[198] Transparency is a key public law principle that has arguably already been jeopardised, alongside other fundamental rights such as the right to a fair trial. The case of Loomis v Wisconsin (as discussed in Chapter One) highlights this issue. Here it was argued

[195] Felzmann, H. Villaronga, E.F. Lutz, C. Tamo-Larrieux, A., 'Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns' (2019) *Big Data & Society*. 1–14

[196] Hayes, E. Wall, S. 'The legal risks of automated decision-making' (peoplemanagement.co.uk, 2020) *People Management CIPD* <https://www.peoplemanagement.co.uk/experts/legal/the-legal-risks-of-automated-decision-making> accessed 23/11/2022

[197] Ibid

[198] Finck, M., 'Automated Decision-Making and Administrative Law' (2020) *Max Planck Institute for Innovation & Competition Research Paper No. 19-10* <https://ssrn.com/abstract=3433684>

that an algorithm that could calculate the likelihood of recidivism used in the sentencing of Loomis was shrouded in secrecy and violated Loomis' right to due process.[199]

This means that with regard to algorithmic decision-making, transparency is now a key concern, and a lack of transparency usually aids occurrence of algorithmic bias.[200] If the system is not transparent and the outputs are discriminatory, it is going to be more difficult to find the source of the bias. As a result, it would appear that the primary focus in creating more responsible AI should be on creating more transparent systems. According to Felzmann et al.,[201] this transparency can take two forms: either prospective or retrospective. The former provides information about how the system operates from the start, whereas the latter describes how a decision was reached after the process is complete, providing a retrospective explanation for the process.[202] Arguably, both of these approaches could be used in order to help uncover the bias that exists within many algorithmic decision-making processes.

Despite this wish for opacity often by those affected, there are also concerns that the desire to achieve such a high standard of transparency within intelligent systems is unrealistic.[203] This is primarily because automated decision-making is often compared to human decision-making; with regard to the latter, a logical explanation is typically achievable, whereas for the former this is not always the case. The aim of many contemporary pieces of legislation and recent amendments to such is to increase transparency in AI-based systems. As a result, it is interesting to consider how effective these current legal frameworks are in safeguarding against and preventing discrimination within AI-based systems.

2.6 Which legal safeguards might help to tackle AI-based discrimination?

With the exponential growth of AI and the evolution of an 'algorithmic society', it is integral that current legislation and regulatory frameworks have either adapted to or are able to adapt to accommodate these technological advancements.[204] As discussed previously, it is now widely acknowledged that there is a bias and discrimination problem within AI. It is also

[199] Smith, R. A. 'Opening the lid on criminal sentencing software' (today.duke.edu, 2017) <https://today.duke.edu/2017/07/opening-lid-criminal-sentencing-software> accessed 23/11/2022
[200] Burrell, J., 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) *Big Data & Society*. 1–12
[201] Ibid n195
[202] Ibid
[203] Zerilli, J. Knott, A. Maclauri, J. Gavaghan, C., Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? (2019) *Philosophy & Technology.* 32, 661–683
[204] The Community Research Development Information Service (CORDIS), 'Safeguarding Equality in the European Algorithmic Society: Tackling Discrimination in Algorithmic Profiling through EU Equality Law' (cordis.europa.eu, 2020) <https://cordis.europa.eu/project/id/898937> accessed 23/11/2022

now evident that there is a universal understanding that we must have legal safeguards in place that adequately protect us from these risks.

Therefore, it is important to examine the current legal frameworks already in place in order to assess their success in dealing with AI-based discrimination and to identify any potential areas for improvement. This section considers the effectiveness of GDPR, in particular the rules on automated decision-making included within Article 22 in tackling AI-based discrimination. This section also considers relevant anti-discrimination regulations such as the ECHR and other supplementary legislations like the Equality Act 2010, which functions as the primary source of anti-discrimination law in the UK. Much of the literature on the effectiveness of current legal safeguards in dealing with the issues presented by AI, although there is little, does so from a distinctly EU perspective.[205] This work, however, assesses not only EU-based regulation but also legislation domestic to the UK.

2.6.1 GDPR

It is customarily acknowledged that protection of personal data is a fundamental right and as such must be upheld.[206] From a European perspective, from around 2010, there has been acknowledgement that existing data protection laws were no longer adequate in meeting the challenges presented by new technological developments.[207] In 2012, the European Commission expressed their position that there was a need to make privacy rights, with regard to use of data online, more robust and reform of the existing Data Protection Directive (Directive 95/46/EC) was necessary.[208] With the implementation of GDPR, it was intended that the use and processing of personal data, particularly when processed in an automated fashion, would be effectively dealt with under the new regulations.

Interestingly, and as highlighted by Drożdż,[209] it is important to note the difference between the old directive and the new regulations; where an EU directive leaves some discretion as to how principles are to be incorporated domestically, the creation of a regulation ensures 'homogenous' implementation of the ruling principles in each applicable state. Arguably, this ensures more stringent data protection rules that are more secure in their protection of

[205] Borgesius, F.J.Z. 'Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence' (2020) *The International Journal of Human Rights.* <https://www.tandfonline.com/doi/full/10.1080/13642987.2020.1743976?scroll=top&needAccess=true > accessed 23/11/2022
[206] Ibid n25, Article 1
[207] European Data Protection Supervisor, 'The History of General Data Protection Regulation' (edps.europa.eu, 2018) <https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en> accessed 23/11/2022
[208] Ibid
[209] Drożdż, A, 'Protection of Natural Persons with Regard to Automated Individual Decision-Making in the GDPR' (2020, Netherlands: Kluwer Law International B.V.)

personal data use. Therefore, from the outset it is clear that the implementation of GDPR was intended to be binding across jurisdictions and to promote trust within those whose data would be processed and used particularly in an automated setting. However, with the rapid development of intelligent technology, and the ability for devices to now make inferences about one's personal attributes with ease, it is questionable as to whether this trust can be upheld, and if it even existed in the first place.

Enshrined in UK law via the Data Protection Act 2018, the scope of the GDPR is stated as applying to the processing and use of personal data, whether by fully or semi-automated means, or where data are processed in a different manner.[210] Here, we can see efforts to break into the 'black box', with efforts being made to ensure that regulations exist to govern the use of personal data within automated decision-making processes. There are also references throughout the regulations to transparency, particularly with regard to information and communication between the controller and the data subject.[211] This mention of transparency is encouraging; it signals that there is a growing understanding of the importance of transparency within AI-based systems.

With regard to the rights of the data subject as listed under Chapter 3 of the GDPR, including Articles 12–23, data subjects have the rights to receive information about the use of their personal data, to access the data and to have this provided in an easily accessible format, to rectify incorrect personal data, to be forgotten, and to restrict use of their data. Interestingly however, with regard to the rights of the data subject as found within Chapter 3 of the Regulations, there are no specific rules governing the rights that an individual has regarding assumptions made about them via an automated system.[212] These are the type of assumptions discussed earlier in this chapter, the type of inferences that often stem from bias and lead to discrimination. For a closer look at the rules regulating automated decision-making, it is necessary to consult Article 22 of the GDPR.

Article 22

From the outset, it is clear that one of the primary purposes of Article 22 is to prohibit discrimination from occurring within an automated decision-making setting. This is evident in that Article 22 clearly states that decisions made regarding an individual should not be based upon any factor included within the 'special category of personal data',[213] i.e. race, religion, and sexuality.[214] Reference to profiling is also made in Article 22, whereby it is listed that a

[210] Ibid n25 Art. 2
[211] Ibid n25 Art. 12
[212] Ibid n25 Chapter 3
[213] Ibid n25 Art. 22, Para. 4
[214] Ibid n25 Art. 9, Para. 1

data subject also has the right to not be subject to a decision based solely on profiling;[215] this typically includes using personal data in order to analyse, predict, or make inferences regarding an individual's probable behaviour and abilities.[216]

Initially, it would appear that all bases are covered by GDPR with regard to the processing of personal information by automated means, and in turn the occurrence of discrimination is minimised. However, upon closer analysis, it can be seen that there are a number of issues. First, and most importantly, it would appear that in theory removing protected characteristics such as ethnicity or sexual orientation from a data set and then allowing decisions to be made based upon the remaining data would be enough to satisfy the regulations as per Article 22. The decision reached via the automated system would not be based upon a special category of personal data but on the other 'non-special' data in the data set.[217]

Yet we know that removing data pertaining to a protected characteristic does not mean that the algorithm is incapable of discrimination. As already discussed with regard to the Apple credit card, removing sensitive data relating to a protected characteristic such as religious beliefs or ethnicity does not mean that bias and discrimination are avoided: it can in fact exacerbate the issue.

Although a protected characteristic such as ethnicity might be deliberately excluded from a data set, inferences can still be made regarding an individual's ethnic background based upon other arguably 'non-special' data that remain, for example, their residential postcode.[218] Therefore, removing the offending data (relating to a protected characteristic) is not necessarily enough to prevent discrimination from occurring. It is therefore believed that these inferences, and lack of robust legal mechanisms governing them, are a good reason to reform data protection law and that this would better protect individuals from AI-based discrimination.[219]

This gives rise to a second issue. Inferences made by an automated system through processes such as profiling are typically not classed as 'special' data and so do not fall

---

[215] Ibid n25 Art. 22, Para. 1

[216] Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/670 Adopted on 3 October 2017 As Last Revised and Adopted on 6 February 2018.* (ec.europea.eu, 2017) <http://ec.europa.eu/justice/data-protection/index_en.htm> accessed 23/11/2022

[217] Baldini, D., *Article 22 GDPR and prohibition of discrimination. An outdated provision?* (cyberlaws.it, 2019) <https://www.cyberlaws.it/2019/article-22-gdpr-and-prohibition-of-discrimination-an-outdated-provision> accessed 23/11/2022

[218] Ibid

[219] Tene, O. Polonetsky, J., 'Big Data for All: Privacy and User Control in the Age of Analytics' (2013) *Northwestern Journal of Technology and Intellectual Property*. 11(5), 239–273, Wachter, S. Mittelstadt, B., 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) *Columbia Business Law Review.* 2019(2), 494–620

within one of the categories as listed within Article 9 of the GDPR.[220] As per Wachter and Mittelstadt[221] and the Article 29 Data Protection Working Party,[222] there is a case to be made that these inferences create a 'new' type of personal data that if rendered identifiable would fall within this 'special' category as per Article 9. With this in mind, it is evident that there are obvious gaps within the protections awarded by GDPR; although the regulations aim to decrease occurrences of discrimination and bias within automated, intelligent systems, the regulations are not completely effective.

Thus, the efforts made via GDPR to introduce a more rigid and uniform set of regulations to govern personal data use online and within automated processes were necessary and to some extent sufficient. It is one of the first wide-scale regulations to directly deal with the issues posed by automated decision-making. Yet GDPR alone does not seem to be the solution to tackle AI-based discrimination. However, as established here, there are accountability gaps with particular regard to both the approach to ridding bias from an automated system and in the approach to which data are and which data should be classified as being 'special' personal data.

Without more refined regulation here, there is a real risk that instances of bias and discrimination will occur and that they will fall outside the remit of GDPR due to the narrow understanding of 'special' data. There are accountability gaps within the regulations, and thus there is room for reform. It is therefore useful to consider which other legal safeguards may be of use in tackling the discrimination and bias problem within AI.*2.6.2 Anti-discrimination laws*

When considering the effectiveness of legal safeguards in tackling discrimination within AI-based systems, it is essential to consider not only relevant data protection regulations but also existing anti-discrimination legislation. The challenges posed by algorithmic discrimination are not limited to data protection and privacy issues, but as is clear, these challenges also include bias-driven unequal treatment of those within our society. It is therefore necessary to consider relevant anti-discrimination law and its effectiveness in dealing with AI-based discrimination. Thus, it is of critical significance to consider how widely current legislation would have to be interpreted in order for it to apply to instances of algorithmic discrimination and if this wide-scale interpretation is possible.

It is believed that a rather fragmented approach to anti-discrimination law is evident across a variety of jurisdictions. This means that some states have legislation that is better equipped

[220] Ibid
[221] Ibid
[222] Ibid n216

than others to deal with the bias and discrimination problem present within AI.[223] This is notably the case with regard to US anti-discrimination law, which has been described as falling short in safeguarding against AI-based discrimination when compared to anti-discrimination regimes in place in areas such as Europe.[224] This is in the sense that the Podesta Report released by the White House in 2014 recommended that US enforcement agencies should find new ways of interpreting existing law, yet existing laws failed to acknowledge or recognise many of the issues stemming from data mining or other AI-related activities in the way that some European initiatives do, even if to a limited extent (such as GDPR).[225]

2.6.2 Anti-discrimination laws

In a similar vein to the discussion regarding data protection, it is worthwhile considering anti-discrimination safeguards present within Europe that may help in tackling AI-based discrimination. As such, a useful place to start is with both the ECHR and the Charter of Fundamental Rights of the European Union (the Charter). The former binds all of its signatories and is enshrined in UK law via the Human Rights Act 1998. The latter applies to EU member states particularly when implementing EU law. Both of these legal safeguards contain provisions that prohibit discrimination. Article 14 of the ECHR, and as a result the Human Rights Act 1998, distinctly prohibits discrimination on a series of protected characteristics similar to those listed within the GDPR; these include:

sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.[226]

And likewise, Title III of the Charter (specifically Article 21) forbids discrimination based upon a similar list of protected characteristics.[227]

Within the UK, it is worthwhile considering the Equality Act 2010 in particular. This piece of legislation is slightly different in that it offers more general protection against discrimination within various aspects of daily life.[228] This is in comparison with that of the Human Rights Act 1998; this Act's focus is on discrimination affecting one's enjoyment of a human right.[229]

---

[223] Hacker, P., 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law' (*Common Market Law Review*. 55, 1143–1186
[224] Barocas, S. Selbst, A.D., 'Big Data's Disparate Impact' (2016) *California Law Review*. 104(3), 671–732
[225] Ibid
[226] European Convention on Human Rights, Art. 14
[227] Charter of Fundamental Rights of the European Union 2012/C 326/02, Title III, Art. 21
[228] Equality and Human Rights Commission 'Article 14: Protection from Discrimination' (equalityhumanrights.com, 2018) <https://www.equalityhumanrights.com/en/human-rights-act/article-14-protection-discrimination> accessed 23/11/2022
[229] Ibid

Despite the abundance of anti-discrimination provisions in place, particularly within the UK, it is suggested that at present these provisions are equipped to tackle algorithmic bias to a certain extent, but they are not fully competent and inclusive as of now.[230]

Arguably, one of the primary issues with regard to the effectiveness of anti-discrimination legislation in tackling AI-based discrimination lies within the very nature of automated decision-making and profiling. Mann and Matzner contend that the whole point of profiling, predicting, and inferring is to find pieces of information that are not directly provided by a given individual; the purpose of this process being to use this new information to find "differences among people to entail that they are treated differently".[231] This is in contention with the entire purpose of anti-discrimination law, which is to prevent individuals from being treated differently on the basis of their differences.[232]

Therefore, it would appear that the essence of algorithmic decision-making, and particularly predictive profiling, is at odds with anti-discrimination principles. With this theory in mind, it is highly unlikely that we would be able to find any current anti-discrimination legislation that fully safeguards against algorithmic discrimination without a specific provision included to deal with inferences and predictions made by automated systems.

This issue regarding inferences is similar to that which arises in relation to data protection laws. While the use of inferences and predictions to classify people based upon their differences is seemingly at odds with anti-discrimination principles, a combination of factors that are not strictly classed as a protected characteristic when combined can still result in discrimination and cause disadvantage. This could be the case despite the fact that these factors (again, such as online shopping habits and holidaying choices) are not classed as protected characteristics capable of causing discrimination. As a result, a number of revealing factors, such as a person's postcode, are typically not deemed to be capable of causing discrimination when in actual fact if combined they are. Therefore, it appears as if this is a commonality that exists between both data protection regulations and anti-discrimination measures.

Consequentially, it would appear that current legal safeguards as discussed in this section, including GDPR, the ECHR, and the Equality Act 2010, are to a very limited extent capable of tackling AI-based discrimination. In order for existing legislation to effectively deal with the causes of algorithmic bias and discrimination as discussed in this work, significant reform is

[230] Ibid n157
[231] Mann, M. Matzner, M., 'Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination' (2019) *Big Data & Society,* p. 4 <https://journals.sagepub.com/doi/10.1177/2053951719895805>
[232] Ibid

necessary. The primary change would appear to be the need to recognise certain factors as being capable of causing discrimination despite not being strictly classed as protected 'special' characteristics. Another potential alternative would be a more innovative approach, to create a new type of legal safeguard that encompasses and combines both data protection law and anti-discrimination principles. A combination of these two types of legislation would be arguably better equipped to tackle AI-based discrimination than just one on its own; further details with regard to this will be included within the final section of this chapter.

2.7 Case analysis

Following an analysis of existing legal safeguards and their effectiveness in tackling AI-based discrimination, it is worthwhile considering these safeguards within the context of current litigation. In particular, this section considers the recent action brought by The Joint Council for the Welfare of Immigrants and 'tech-justice' group Foxglove against the Home Office regarding the use of an algorithm within the Home Office's visa application process.[233] The algorithm in question used is what can be described as a 'traffic-light system' for rating visa applicants. The issue here appeared to be that the algorithm rated and discriminated against individuals based on their nationality.[234] Individuals from 'higher risk' nations were given a lower rating than their counterparts from more 'suitable' nations, which meant that their applications were scrutinised in much more depth, prolonging the process and meaning they would likely be denied a visa.

Interestingly, the legal challenge brought by The Joint Council for the Welfare of Immigrants and Foxglove was founded on the basis that the use of this algorithm directly violated the Equality Act 2010, as it discriminated against applicants based on their race. Their action was successful in that at the beginning of August 2020, the then Home Secretary Priti Patel agreed to cease the use of the visa application system in question and to review the process. Therefore, this gives some indication that current legal safeguards, in particular anti-discrimination laws such as the Equality Act 2010, are to some extent successful in protecting against algorithmic bias.

It is worth noting however that discrimination based upon an individual's nationality, which is a protected characteristic, is clearly recognised as being unlawful. If other 'non-special' characteristics were used in this instance to make inferences and thus a decision about a

[233] The Joint Council for the Welfare of Immigrants 'We Won! Home Office to Stop Using Racist Visa Algorithm' (jcwi.org.uk, 2020) <https://www.jcwi.org.uk/news/we-won-home-office-to-stop-using-racist-visa-algorithm> accessed 23/11/2022
[234] Ibid

person, there is a likelihood that discrimination and unfair treatment will still occur. Unfortunately, this discrimination would likely not be recognised or protected by the Equality Act 2010 and other similar legislations. It is therefore encouraging that current legislation can and has successfully been used to tackle AI-based discrimination in this instance. However, it is evident that there is still some way to go before we have fully competent anti-discrimination safeguards that can tackle algorithmic bias and discrimination.

With regard to this action brought against the Home Office, it was contended that following incidents such as the Windrush scandal, it was obvious that the Home Office had an entrenched history of racism. As such, these historical prejudices against particular nationalities formed the basis of the visa application software in question.[235] Interestingly however, an issue still remains with regard to transparency. We know that the algorithm in question here used a list of suspect nationalities to discriminate against applicants; however, the Home Office would not provide any further information regarding other factors that were considered by the algorithm when reaching a decision.[236] Once again, we see the potential issues posed by a lack of transparency within automated decision-making systems. We know that discrimination was established on the grounds of racism; however, we are unaware of other factors involved in the process that could have caused further discrimination on different bases.

The opinions held by The Joint Office for the Welfare of Immigrants regarding safeguards against algorithmic decision-making are in line with the ideas presented in this chapter. At some point, a combination of anti-discrimination and data protection principles would form the basis of adequate measures to be used in tackling algorithmic discrimination. The Joint Office for the Welfare of Immigrants states that the then Home Secretary Priti Patel agreed to implement their agreed legal measures which come in the form of an Equality Impact Assessment and a Data Protection Impact Assessment. However, this is only in relation to the aforementioned visa application process. Therefore, it is hypothesised that a similar approach taken to similar algorithmic decision-making processes could help tackle existing discrimination caused by algorithms and bias present within the automated process. By employing the approach adopted here, potential instances of discrimination, bias, and unfair treatment are more likely to be identified, which provides more robust grounds for tackling algorithmic discrimination on a wider scale.

This case highlights and re-enforces a number of issues discussed so far within this work. There is a clear issue with regard to historical prejudices being present within data sets used

[235] Ibid, Para. 3
[236] Ibid

to train algorithms meaning that these biases are exacerbated, there is an inherit lack of transparency within most algorithmic decision-making processes, and reform to current legal safeguards is necessary in order to fully tackle the number of issues presented by AI-based discrimination. Unfortunately, with the inevitable continued reliance upon automated decision-making, further legal challenges are also to be expected. However, as with cases like that brought by The Joint Council for the Welfare of Immigrants against the Home Office, it is encouraging that there appears to be further recognition of this bias and discrimination problem within intelligent systems and some limited efforts being taken to tackle the issue at its core.

2.8 Conclusions

This chapter has demonstrated clearly that one of, if not the most pertinent issues posed by AI at present is the inherent bias and discrimination problem. The issues posed by algorithmic bias and discrimination are obvious, and there is clear evidence that we are becoming increasingly aware of the potential wide-ranging impacts that these issues have. As demonstrated in this case study, the first line of defence in tackling discrimination, particularly within automated decision-making systems, is via legal measures such as legislation, regulation, and policy. Resultantly, it would appear that a key focus going forward in the response to these tech-based issues would be to more closely consider the effectiveness and functionality of the legal safeguards that may be used to tackle occurrence of bias and discrimination within AI-based systems. This analysis would feature scrutiny of not only domestic law but also international law.

As is present in the evaluation provided in this chapter, it is clear that there is room for improving the effectiveness of existing legal safeguards such as, but not limited to, GDPR and the Equality Act 2010 to fully protect against the risk of discrimination present within these automated decision-making systems. There are a number of ways in which this could be achieved; one of which would be by establishing a set of measures that incorporate both personal data protection and anti-discrimination principles as suggested by The Joint Council for the Welfare of Immigrants in their attempt to reform the current visa application process within the UK. This reform to the law could take the shape of a requirement for organisations intending to deploy AI to have to conduct a combined equality impact assessment and data protection impact assessment prior to its use, as suggested by The Joint Council for the Welfare of Immigrants.

In addition to this and as highlighted throughout this chapter, efforts could and should be made to recognise certain pieces of ‘non-special’ information (such as a person’s postcode) as protected and categorically ‘special’ data. This is primarily because of the capability

available to use these data in order to make inferences about an individual that can once more reveal a protected characteristic and lead to discrimination. Recognition of this specific data type could be incorporated within legislation such as GDPR and the Equality Act 2010 as discussed already in this chapter, alongside other legislation that specifically references already recognised protected characteristics, such as the Human Rights Act 1998.

As a result, the second research objective of this thesis, which was to examine how well equipped current legal instruments are in dealing with the issues posed by AI, has been addressed in this chapter. This chapter proposes that reform to the laws discussed herein are necessary in order to properly safeguard against the risks of discrimination posed by the increasing use of automated decision-making systems, and as such reasonable recommendations are offered as the fourth research question sets out to deliver.

Continuous and widespread use of AI and other intelligent systems in order to assist with everyday tasks is inevitable: any attempt to stop this is counterproductive and futile. Therefore, it is recommended that one part of a multifaceted approach in dealing with this problem is to begin by modifying the aforementioned laws in such a way that it is more suited to adequately protect against the risks posed by these automated systems.

Similarly, the issues raised within this chapter provide further support for the regulatory proposals discussed further on in this thesis.[237] In addition to the modification of existing law, additional regulatory measures will be necessary to ensure that the bias and discrimination problem with AI is eliminated at the source, therefore limiting the likelihood that this will further affect society. With this in mind, it is worthwhile considering the approaches being taken by various governments and organisations across the world to tackle the issues AI pose at large, and whether or not these approaches are effective and functional.

[237] See Chapters Five and Six

Chapter Three

Comparing National Strategies and Frameworks on AI: The UK, US, China, South Africa, and Egypt

3.1 Introduction

The exponential growth of artificial intelligence, the evolution of an 'algorithmic society'[238], and the established problem regarding algorithmic bias and discrimination (used as a case study in the previous chapter) serves to demonstrate that regulation within this space is necessary; whether this includes reforming existing, outdated legal measures or creating something altogether new. We have acknowledged that there is a bias problem within AI, but there are issues far beyond this, AI also poses threats to our democracy for example. Therefore, it is evident that there must be adequate legal safeguards put in place to protect against the risks posed by artificially intelligent systems. The question remains as to what shape they will take, however.

It is interesting to note the various and somewhat segregated approaches that have, and are, currently being taken to regulate AI. For example, as opposed to focusing on a baseline, catch-all approach, academics, and lawmakers are focusing efforts on regulating specific applications of AI. An example of this, and one that has received much consideration is the regulation of artificial intelligence used in armed conflict.[239] The use of AI in this manner has the potential to have considerable and potentially lethal impacts, yet the impact of using AI in everyday functions such as within banking or healthcare also has the potential to negatively affect society in a multitude of ways.

Although several nations and regions appear to agree upon some factors and principles that should be included within regulation for AI such as accountability, transparency, and non-discrimination, they either fail to offer a reasonable solution for achieving these goals, or advocate for considerably different ways of accomplishing them.[240] Similarly, there are proposals to make research on artificial intelligence and its potential capabilities immune from the application of any future AI regulations and legislation,[241] meanwhile others have made the case that research on dangerous, potentially lethal AI should be banned

[238] The Community Research Development Information Service (CORDIS) 'Safeguarding Equality in the European Algorithmic Society: Tackling Discrimination in Algorithmic Profiling through EU Equality Law' (cordis.europa.eu, 2020) <https://cordis.europa.eu/project/id/898937> accessed 23/11/2022
[239] D. Lewis, 'International Legal Regulation of the Employment of Artificial-Intelligence-Related Technologies in Armed Conflict' (2020) Moscow Journal of International Law 2 53-64
[240] M. C. Buiten, 'Towards Intelligent Regulation of Artificial Intelligence' (2019) European Journal of Risk Regulation 10 41-59
[241] R. Calo, 'Open robotics', (2011) Maryland Law Review 70(3) 101-142.

completely.[242] All the while there is the general agreement that any regulation should be future proofed, flexible and amenable to change.[243] Thus it is only too clear that there is conflict with regard to the most effective way to regulate AI, and as a result an analysis of these competing viewpoints is valuable in order to consider what the most agreeable, unobtrusive approach to regulation might be.

It is equally important to examine the current legal frameworks some states may already have in place in order to both determine adequacy of present safeguards in dealing with AI-related risks such as the general use of autonomous vehicles, and to identify potential areas and scope for improvement. This section of the thesis considers the effectiveness of the regulatory strategies and frameworks for AI by a variety of nations across the world including the following; approaches taken by the UK in its National AI Strategy,[244] China's AI development plan,[245] the more situation-specific approach adopted by the US,[246] the approach considered in South Africa,[247] and Egypt's National AI Strategy.[248] These nations have been chosen in particular as they represent countries in both the global north and global south, all of which have differing interests in AI development and therefore share some unique differences in their preparedness and reasons for regulating AI.

Comparing the approaches taken by each of these nations is necessary in order to for us to properly understand how aligned proposed legal frameworks and strategies are across the globe, as well as understanding how perspectives on AI use, development and regulation differ across nations. It will allow us to better judge how successful particular regulatory approaches might be in different areas of the globe. The analysis contained in this Chapter and the following Chapter will also help to identify how each of the selected nations, regions

---

[242] Future of Life, 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers' (*futureoflife.org,* 2015) <http://futureoflife.org/open-letter-autonomous-weapons/> accessed 10/07/2021

[243] N. Petit, 'Law and Regulation of Artificial Intelligence and Robots – Conceptual Framework and Normative Implications' (2017) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339> accessed 10/11/2021

[244] Office for AI, Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, 'National AI Strategy' (2021) <https://www.gov.uk/government/publications/national-ai-strategy> 10/11/2021

[245] H. Roberts, J. Cowls, J Morely, M. Taddeo, V. Wang, L. Floridi, 'The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation' (2021) *AI & Society* 36, 59-77

[246] National Conference of State Legislatures, 'Legislation Related to Artificial Intelligence' (ncsl.org, 2022) <https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx#2021> accessed 10/11/2021

[247] Department of Telecommunications and Postal Services, 'Presidential Commission on Fourth Industrial Revolution' (oecd.ai, 2019) < https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26873> accessed 26/03/23

[248] The National Council of Artificial Intelligence, 'Egypt National Artificial Intelligence Strategy' (mcit.gov, 2022) < https://mcit.gov.eg/Upcont/Documents/Publications_672021000_Egypt-National-AI-Strategy-English.pdf> accessed 26/03/2023

and organisations within this thesis are aiming to tackle the most common ethical issues we see arising from AI use, namely accountability, transparency, and non-discrimination.

Therefore, the following sections will examine the general AI strategies and regulatory frameworks set out in each of the chosen nations. This will be done with a view to considering the general structure and aims of the selected frameworks and strategies, identifying any strengths and shortcomings of the approaches, as well as any potential suggestions for improvements to these regimes and proposals where appropriate. It is also key to point out that there are a mix of enforceable legislative efforts and strategies examined in this chapter and the following chapter, for example the UK strategy examined here merely demonstrates intentions in this space, whereas some of the US regulations considered are already legally enforceable, therefore giving a good overview of the current state of global preparedness for AI regulation.

3.2 Regulatory strategies in the UK

The UK government recognises the impact that AI can potentially have upon the public, and as such one particular area of focus for the government has been considering more closely the responsibility and duty owed to the public when AI is used within the Public Sector.[249] Further recommendations made by The Committee on Standards in Public Life (which are in line with those made by this thesis) suggest the need for ensuring that use of AI by public bodies is done in line with anti-discrimination law, that the role of impact assessments are considered, and that there should be harmonisation of current ethical principles and guidance.[250]

This particular focus upon regulating the use of AI by public bodies comes down to the simple notion that individuals can decide against using the services that a private sector body provides, but often cannot opt-out of the services provided by a public sector body.[251] This is a valuable point and provides good rationale for the need to regulate AI, however, the importance of regulating AI in such a way that includes the private sector within its scope is also important. Although there appears to be a dilemma in that the way in which AI is

[249] The Committee on Standards in Public Life, 'Artificial Intelligence and Public Standards' (GOV.uk, 2020) p. 12 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF> accessed 10/11/2021
[250] ibid
[251] ibid

regulated within the private sector is currently via voluntary, internationally agreed standards or principles (such as the OECD Principles on AI)[252] which are not legally binding.[253]

However, creating regulation that finds a balance between business incentives, investment in the development of AI, ensures the safety of people and also results in penalties for non-compliance, is tough to strike. This section of the chapter considers regulatory approaches taken by the UK, critically analyses these approaches and compares them with those taken by the other 'big tech players'.

3.2.1 The National AI Strategy

The most logical place to begin is by considering the UK's National AI Strategy (referred to as the Strategy) published in September 2021,[254] which was further supported via the White Paper on the same topic published by the Department for Science, Innovation and Technology, and the Office for Artificial Intelligence in Spring 2023.[255] An official stance on the UK's potential AI strategy has been a long time coming, and so the document definitely makes for an interesting read. Upon first glance, it is clear that the intentions of the UK government are to use this particular strategy to confirm the UK as a 'global AI superpower', with a considerable fifty pages out of a sixty-two-page document focused upon promoting and developing the UK's AI ecosystem.[256]

This is particularly interesting as it consolidates the idea that the UK government are approaching their AI policies from a distinctly internationally collaborative perspective.[257] Specifically, Nadine Dorries' (Secretary of State for Department of Digital, Culture, Media and Sport at the time) stated:

"This National AI Strategy will signal to the world our intention to build the most pro-innovation regulatory environment in the world"[258]

This statement alone raises a number of questions, primarily, how exactly do the UK government intend to position themselves as 'the most pro-innovation regulatory environment'? This potentially implies that there may be scope to amend and even reduce

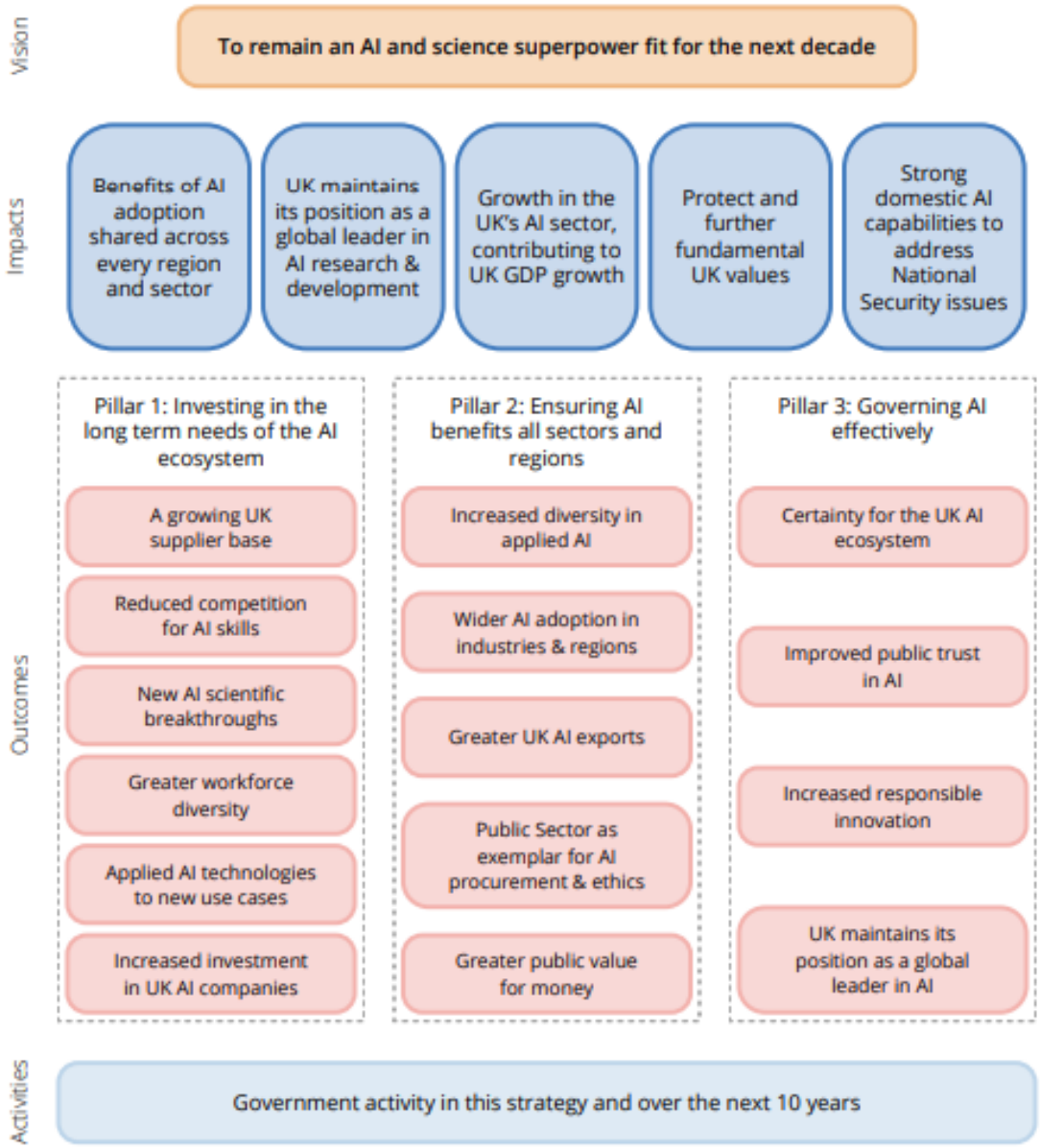
[252] OECD, 'OECD Council Recommendation on Artificial Intelligence' (oecd.org, 2019) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 10/11/2021
[253] Elliot Wellsteed-Crook, 'Regulate tech to realise the benefits' (newstatesman.com, 2020) <https://www.newstatesman.com/spotlight/emerging-technologies/2020/09/regulate-tech-realise-benefits> accessed 12/11/2021
[254] Ibid n215
[255] Department for Science, Innovation and Technology, Office for Artificial Intelligence, 'A pro-innovation approach to AI regulation' (gov.uk, 2023) < https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper> accessed 05/04/2023
[256] Ibid n244
[257] Ibid n244 p. 5
[258] Ibid n244 p.5

the scope of application of certain policies (such as data protection policies that are in line with the notoriously strict GDPR for example) in order to align with other global partners, especially since the UK's departure from the EU. The implications of this manoeuvre however are in need of consideration and will be considered further on in this chapter.

It is worth noting that this particular document, despite the numerous proposals for transformation, should realistically read as a 'signalling document' and not legally binding.[259] Yet, despite the recommendatory nature of this text it shouldn't be undervalued; it really does act as an indicator that the relevant policy, research, industry and governance bodies within the UK see AI as worthy of singular consideration, as opposed to be paired with or forming a part of another government-lead initiative, and gives us a good indication as to the future of AI regulation in the UK. This is promising and will definitely help to pave the path to a collaborative and functional regulatory regime; whether that will be a short-term or long-term goal is yet to be established.

[259] E. Kazim, D. Almeida, N. Kingsman, C. Kerrigan, A. Koshiyama, E. Lomas, A. Hilliard, 'Innovation and opportunity: review of the UK's national AI strategy' (2021) *Discover Artificial Intelligence* 1(14)

Figure 1: The UK's National AI Strategy[260]



Source: Office for AI, Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, 'National AI Strategy' (2021) https://www.gov.uk/government/publications/national-ai-strategy, accessed 10/11/2021

As demonstrated by the above figure, the National AI Strategy is structured in three pillars. Pillar one details the intention to invest in the long-term needs of the AI ecosystem, pillar two pertains to ensuring that AI benefits all sectors and regions within the UK, whilst the third and final pillar lists the intention to govern AI effectively (it is worth noting that this is the least detailed pillar of the three). All of the outcomes listed within these pillars are intended to be

[260] Ibid n244 p. 14

achieved within a 10-year period. Therefore, the following sections will consider these three pillars, their intended aims and outcomes, achievability within the allotted time period and overall likely impact upon the potential AI regulatory environment.

3.2.1.1 Pillar One: Investing in the long-term needs of the AI ecosystem

Pillar one is a crucial step in the overall National AI Strategy, and therefore lays the foundations that will allow the UK to 'retain' (or establish) its AI superpower status.[261] When considering the geopolitical turbulence faced by the UK in recent years (Brexit in particular springs to mind), it is quite clear as to why establishing oneself as a force within the AI-sphere is a predominant interest of the UK government. As the UK transitions out of the European Union, it will likely want to align itself within other like-minded administrations, and more importantly attract business and trade from non-European companies invested in this space. Therefore, the UK is understandably aiming to cement itself in place as a notable AI superpower, and in parallel with the likes of China and the USA.[262]

Research

With the intentions of this pillar and the overall AI strategy clear, it is essential to consider how exactly the UK government intends to achieve this arduous goal. It would appear that the central focus and method of achieving the above objective is by developing skills and talent within the UK's AI habitat, which is in line with some of the proposals made within this thesis.[263] Ensuring that facilities are in place to establish, develop and grow a skilled AI workforce within the UK achieves that very goal of investing in the long-term needs of the AI ecosystem; it provides good grounds for companies and organisations invested in this space to set-up shop in the UK.

There has been considerable effort made by the UK government in this space in recent years, all leading up to this particular point. For example, the Department for Digital, Culture, Media and Sport (DCMS) published their '10 Tech Priorities', with number three on this list pertaining to build a 'tech-savvy nation'.[264] Similarly, in a study carried out by Ipsos Mori in 2020 on the UK AI labour market, it was established that despite a significant number of cyber and AI skills initiatives spearheaded by the government there was still a significant

[261] Ibid n244 p. 22
[262] K. F. Lee, *AI Superpowers: China, Silicon Valley and the New World Order* (Houghton Mifflin Harcourt, New York, 2018)
[263] See Chapter Five for further detail on proposals regarding improving education and AI skills within the proposed regulatory strategy
[264] Department for Digital Culture, Media and Sport, 'Our 10 Tech Priorities' (*DCMS.gov.uk*) <https://dcms.shorthandstories.com/Our-Ten-Tech-Priorities/index.html> accessed 02/12/2021

shortage in the number of workers with sufficient AI skills, yet the demand for these workers is increasing.[265]

As a result, significant focus here is placed upon supporting institutions to nurture those engaged with AI at an academic level, for example by helping individuals to pursue postgraduate courses in AI and to keep individuals who possess the desired knowledge and skill sets working within this space.[266] The UK are not the only country concerned about research output. During the past two decades the US topped the charts for having accumulated the highest number of AI publications deriving from US-based institutions and organisations, yet China overtook this publishing record by producing more AI-related papers than any other country between 2016 to 2019, publishing around 30,000 more papers than the US in 2019 alone.[267] These statistics alone demonstrate the real race that exists between these nations in their hopes to become the leading authority in all things AI.

Yet, the quality of these papers is debatable, with the robustness and academic integrity of some Chinese publications specifically in question. During 2019, Chinese AI papers were cited 20% less than the world average, whilst US AI papers were cited 40% more than world average.[268] It is interesting to consider this in line with the decision reached by Chinese government in early 2020 to ban cash reward incentives for publishing papers in hopes of promoting production of further high impact research.[269] As per the 2021 AI Index Report, the competition is very much still on between the US and China with China taking the lead in AI journal publications and the US taking the significant lead in AI conference publications.[270]

---

[265] Ipsos Mori, 'Understanding the UK AI labour market: 2020' (*GOV.uk,* 2020) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984671/DCMS_and_Ipsos_MORI_Understanding_the_AI_Labour_Market_2020_Full_Report.pdf> accessed 02/12/2021

[266] Department for Business, Energy, & Industrial Strategy, Department for Digital, Culture, Media and Sport, Office for Artificial Intelligence, 'Turing Artificial Intelligence Fellowships' (*GOV.uk* 2021) <https://www.gov.uk/government/publications/turing-artificial-intelligence-fellowships/turing-artificial-intelligence-fellowships> accessed 02/12/2021
For example, in July 2021 the UK government announced that they would be helping to fund the Turing AI Fellowship scheme costing around £46 million. In specific, the scheme focuses on 'retaining, attracting and developing' researchers in this space.

[267] N. Savage, 'The race to the top among the world's leaders in artificial intelligence' (2020) *Nature Index* <https://www.nature.com/articles/d41586-020-03409-8> accessed 03/12/2021

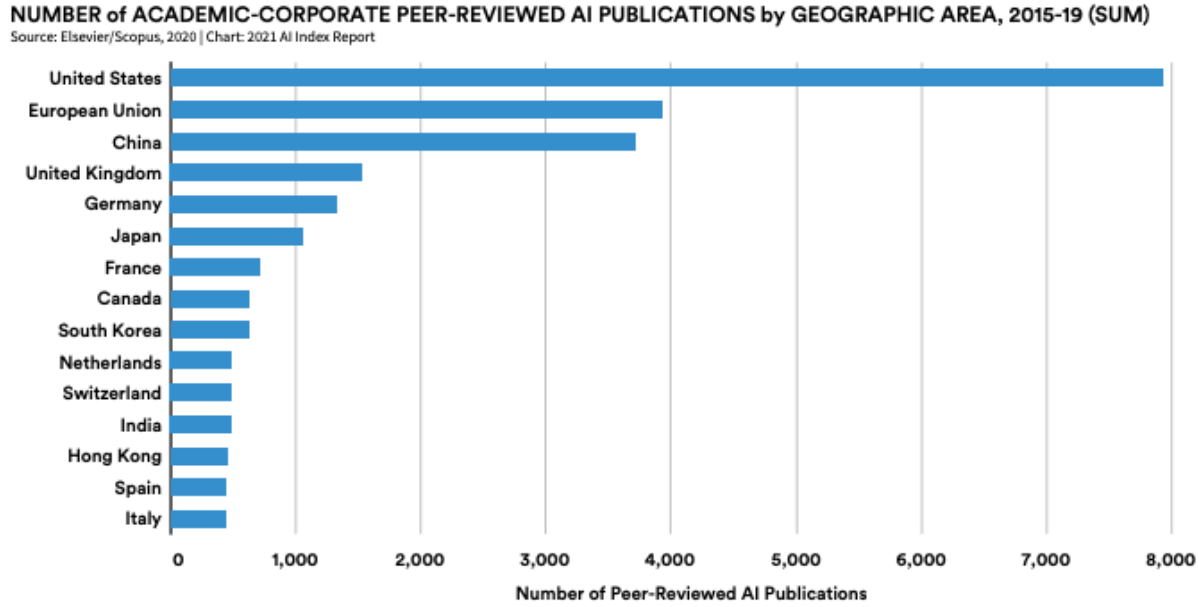
[268] Human-Centered Artificial Intelligence at Stanford University, 'Artificial Intelligence Index Report 2019' (hai.stanford.edu, 2019) <https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf> accessed 03/12/2021

[269] S. Mallapaty, 'China bans cash rewards for publishing papers' (nature.com, 2020) <https://www.nature.com/articles/d41586-020-00574-8> accessed 03/12/2021

[270] Human-Centered Artificial Intelligence at Stanford University, 'Artificial Intelligence Index Report 2021' (2021) p. 17 <https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf> accessed 10/12/2021

With such a fierce race existing between two nations of considerable size, two nations that are already well-established within this field, where is the UK placed within this global race? The below figure reveals the deficit that truly exists.

Figure 2: Number of Academic-Corporate Peer-Reviewed AI Publications by Geographic Area, 2015-2019[271]



Source: 'Artificial Intelligence Index Report 2021', Human-Centered AI Institute at Stanford University (2021) p. 17 https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf, accessed 10/12/2021

As demonstrated by the above figure two, the UK are considerably behind when compared to academic-corporate peer-reviewed AI publications produced by the US, the EU and China, publishing only around 1,500 papers in comparison to just under 8,000 produced within the US. As a result, the choice to place a good amount of focus and resource upon developing the UK's role in the global AI research space is likely going to be a good investment, and realistically the outcomes of this should start to become apparent within the immediate future.

Robust and well-informed research have the potential to directly (and positively) impact upon government policy, which aim to tackle the issues we currently face due to AI; namely accountability and transparency and non-discrimination.[272] The added funding and incentive put in place via the UK's National AI Strategy will likely go some way to boosting the UK's high impact research output, specifically with regards to government collaboration with

[271] Ibid p. 23
[272] Knowledge Exchange Unit, UK Parliament, 'Research impact on policy' (*GOV.uk* 2021) < https://www.parliament.uk/globalassets/assets/teams/post/research_impact_on_policy_briefing_document_june21.pdf?__cf_chl_managed_tk__=ytKbXGUsLt8lItZ6DbgMLGWLVlIYFzGiukHRoCjpuUQ-1643039911-0-gaNycGzNCeU> accessed 14/12/2021

institutions such as The Turing Institute and The Ada Lovelace Institute. It is therefore anticipated that an increase in high impact research activity within the UK within the coming years will benefit the future regulatory strategies adopted within the country, something that would be in line with this thesis and welcomed by the author. Hopefully these research outputs will help the UK to create functional regulatory measures that ensure key issues such as accountability and transparency are tackled head-on.

Education

Another proposal which accompanies plans to boost research activity and high-impact outputs within pillar one of the Strategy, is to improve the AI-based knowledge and skills of the existing workforce.[273] This proposal lines up well with the priorities discussed within Chapter Five of this thesis, specifically improving AI explainability by improving public understanding of the technology itself. By ensuring that we improve the skills of those not only within the education system (via national curriculum for example) but those within the workplace too, it is more likely that we can increase trust in AI systems, boost development in this space, and position the UK as a country truly invested in the evolution of AI.

The National AI Strategy proposes that this particular aim can be achieved by making use of existing government policies and initiatives, including for example the Skills for Jobs initiative used to promote lifelong learning for opportunity and growth.[274] Other similar initiatives proposed for use in this capacity include Skills Bootcamps, these are free 16-week courses for adults in all states of employment that aim to equip individuals with sector-specific skills (in this case, machine learning and other AI related skill-sets).[275]

Again, this seems like a logical proposal as without a sufficiently skilled workforce, the UK cannot hope to become the AI superpower it wishes to be. For example, a growing number of jobs require AI-related skills, and there are similarly an increasing number of AI-related job postings in the UK; the number of AI-related job postings found online increased by around 3.6 million postings in just five years between 2012 and 2017.[276] Not only this but the apparent threat that AI automation itself poses to workforces across all sectors means that

[273] Ibid n244 p. 27
[274] Department for Education, 'Skills for jobs: lifelong learning for opportunity and growth' (*GOV.uk*, 2021) < https://www.gov.uk/government/publications/skills-for-jobs-lifelong-learning-for-opportunity-and-growth> accessing 14/12/2021
[275] Department for Education, 'National Skills Fund' (*GOV.uk* 2021) <https://www.gov.uk/guidance/national-skills-fund#skills-bootcamps> accessed 14/12/2021
[276] M. Squicciarini, H. Nachtigall, 'Demand for AI skills in jobs: Evidence from online job postings' (oecd.org, 2021) *OECD Science, Technology and Industry Working Papers* < https://www.oecd-ilibrary.org/docserver/3ed32d94-en.pdf?expires=1643041720&id=id&accname=guest&checksum=24A35C03F178E06818B5B093A47BF001> accessed 14/12/2021

digital/cyber skills are more valuable now than ever, and will likely continue on this trajectory; some estimations suggest that within the next five years around 30% of all work activities could become automated.[277] Therefore, having the relevant skillset that allows job mobility, professional development and the ability to progress within the workplace is all dependent on an adequate AI-skills policy.

How effective would a policy like the one suggested within the UK's National AI Strategy be? Again, this is less of a policy proposal and more of a signalling document. The Strategy refers in part to a number of pre-existing government initiatives, such as Skills for Jobs and Skills Bootcamps discussed already, and how these will be used alongside the 'Skills Value Chain'[278], a scheme already piloted by the Department for Education which at present allows manufacturers to address the skills gaps that exist in their sector in order to exploit new technologies.[279]

However, this latter solution is primarily going to be used in order to gain a better understanding of the extent of the skills gaps that exist across sectors, this information will then hopefully help government departments such as the Department for Education formulate a suitable approach. Resultantly, this is rather a long-term project and one that we will likely not see the benefits of for quite some time. It would be beneficial to see a slightly more robust proposal here within the Strategy, one of similar constitution to the proposals made in order to aid research and scholarship in this space, for example the Turing Fellowship Scheme and the establishment of the Advanced Research and Invention Agency (ARIA), as improving skills within the general workforce is arguably of equal importance.[280]

Data

A final remark regarding pillar one of the Strategy is in relation to data, and more specifically data sharing rights. Good data is the key to AI; without it we cannot hope to create safe, reliable and effective AI systems. At present, and as indicated within the Strategy, there is considerable work ongoing within this space in the UK including continuing development of

[277] L. Good, E. Buford, 'Modernizing and Investing in Workforce Development' (Corporation for a Skilled Workforce, 2021) <https://skilledwork.org/wp-content/uploads/2021/03/Modernizing-and-Investing-in-Workforce-Development.pdf> accessed 15/12/2021
[278] Ibid n244 p. 26
[279] Catapult: High Value Manufacturing, 'Manufacturing the Future Workforce' (hvm.catapult.org, 2021) < https://hvm.catapult.org.uk/mtfw/> accessed 15/12/2021
[280] Department for Business, Energy & Industrial Strategy, 'Advanced Research and Invention Agency (ARIA): policy statement' (GOV.uk 2021) < https://www.gov.uk/government/publications/advanced-research-and-invention-agency-aria-statement-of-policy-intent/advanced-research-and-invention-agency-aria-policy-statement> accessed 15/12/2021

the National Data Strategy,[281] a consultation on the future of data protection in the UK,[282] and exploring existing legal methods for data stewardship.[283]

Although, the Strategy does indicate that the UK government are looking to improve and enable data sharing which would be beneficial to the development and training of various AI systems, they are also looking "to permit the collection and processing of sensitive and protected characteristics data" in order to attempt to monitor the bias and discrimination problem present within AI.[284] AI related bias and discrimination is considered in some depth within Chapter Two of this thesis, and here the effectiveness of GDPR in tackling this issue was assessed in some detail. However, considering the proposals made within the Strategy, this would suggest that the UK government are intending to move away from the GDPR-style framework to something slightly more flexible and generous.[285]

It is worth bearing in mind however, that the GDPR has wide-ranging scope. It applies to those outside of the EU if data belonging to EU citizens are being processed by an organisation or institution. It is suggested that being more flexible in regard to processing of protected and special characteristics data would in fact be a step backwards in the UK's data protection regime, and not a step forward as the Strategy suggests. In fact, this thesis proposes that further measures be added to the GDPR, and the UK iteration, the Data Protection Act 2018 (DPA)[286], in order to prevent certain proxy data being utilised specifically within automated decision-making processes and to class this as a form of special data.[287] The proposals made in the Strategy are completely opposed to this suggestion and it does appear as though this would be a rather regressive policy choice.

The obvious (and likely) reasoning for this apparent Atlanticism is possibly an effort on the UK's behalf to appear less stringent with regard to data protection rights in order to appeal to an American market, to establish and build upon relationships with US partners.[288] This is a question that is very much still open for consideration, and it will be enlightening to see

[281] Department for Digital, Culture, Media and Sport, 'National Data Strategy Mission 1 Policy Framework: Unlocking the value of data across the economy' (GOV.uk 2021) < https://www.gov.uk/government/publications/national-data-strategy-mission-1-policy-framework-unlocking-the-value-of-data-across-the-economy/national-data-strategy-mission-1-policy-framework-unlocking-the-value-of-data-across-the-economy> accessed 15/12/2021
[282] Department for Digital, Culture, Media and Sport, 'Data: a new direction' (GOV.uk, 2021) < https://www.gov.uk/government/consultations/data-a-new-direction> accessed 15/12/2021
[283] Ada Lovelace Institute, AI Council, 'Exploring legal mechanisms for data stewardship' (adalovelaceinstitute.org, 2021) < https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/> accessed 15/12/2021
[284] Ibid n244 p. 31
[285] Ibid n244 p. 31
[286] Data Protection Act 2018
[287] See Chapter 2 for further detail on these proposals
[288] Ibid n283 p. 1

perceptions towards a potential change to the data protection rights via the public consultation on the future of data.

3.2.1.2 Pillar Two: Ensuring AI benefits all sectors and regions

Pillar two considers some similar points of action as pillar one, namely supporting AI innovation and ensuring we all benefit from AI regardless of sector or region. More specifically this involves supporting businesses to better use AI in a way that benefits society, encouraging more organisations to invest in and deploy AI-systems within their enterprise, and ensure that AI is used widely for the public benefit (for example, within the Covid-19 pandemic).[289]

AI for Climate Change

One interesting point to consider within this pillar is the intention to align the Strategy with government incentives on climate change, which is a logical and sensible approach to take. Not only this, but the intention here is to utilise AI in the most effective way possible to help tackle the climate crisis.[290] Some hopeful AI deployments in this space include:[291]

- Using machine learning to monitor the environment
- Using AI within the energy sector and to help control its network distribution
- To use data to identify inefficiencies in emissions-heavy industries
- Using AI within atmospheric modelling in order to combat future issues

Using AI as a tool to tackle climate change is a concept that a growing number of academics and practitioners are beginning to consider. AI is believed to be the key to achieving goals such as global net zero by 2050, set out at the likes of COP26; the combination of human and machine intelligence in this space has the potential to solve some of the most pressing problems of our time.[292] Therefore, any government-led policy that acknowledges the merits of AI within this space, and actively promotes the development and deployment of AI in this capacity is not only beneficial but commendable.

[289] Ibid n244 p. 42
[290] D. Rolnik et al, 'Tackling Climate Change with Machine Learning' (2019) *Computers and Society* <*https://arxiv.org/abs/1906.05433*> accessed 20/12/2021
[291] Ibid n244 p. 45
[292] G. Shaddick, 'COP26 and beyond: the crucial role for AI in tackling climate change' (turing.ac.uk, 2021) <https://www.turing.ac.uk/blog/cop26-and-beyond-crucial-role-ai-tackling-climate-change> accessed 20/12/2021

Inclusion of this particular incentive within the Strategy seems like a good choice, and one that other climate-conscious countries around the world will likely adopt. Despite this though, it is worth noting that using AI for this purpose is not a catch-all solution, it does have its downsides. The training of these AI-systems, including the training of neural networks for example, consumes a considerable amount of energy; this is compared to the actual running of the AI itself once functional.[293] As demonstrated by Strubell et al, to train one single AI model can lead to emissions of almost 300,000 kg of CO_2, this is equivalent to the CO_2 produced by five average cars over the course of their lifetimes.[294] This is combined with mining and extraction of raw materials necessary to manufacture these electronic devices which also leads to considerable environmental risks, making the use of AI in this space a real "double-edged sword".[295]

Although, avoiding using AI to tackle climate change is non-sensical as this technology really does have the potential to be ground-breaking, it would appear as though the real task is balancing both interests, the interests of the public with the interests of the environment. It is suggested that as the UK progresses in this proposal, that attention is paid to this niche issue. For example, when developing and training AI-systems intended to tackle climate change, factors such as energy use could be tracked and reported alongside other performance metrics.[296] Even more simply, increasing awareness and acknowledgement within both the research community and within government regarding the ethical issue that energy use in the training of AI systems poses, would be both desirable and achievable if set out within a future iteration of the Strategy.

3.2.1.3 Pillar Three: Governing AI effectively

The final pillar of the National AI Strategy pertains to the governance of AI, this includes promoting the safe and ethical development of AI whilst ensuring that we as a society are guarded against AI-related risks.[297] Again, the UK government are setting their sights high here, and aim to "build the most pro-innovation system for AI governance in the world".[298] Via the Strategy, the UK government clearly set out within this pillar to achieve a number of

[293] M. Coeckelbergh, 'AI for climate: freedom, justice, and other ethical and political challenges' (2021) *AI & Ethics* 1, 67-72
[294] E. Strubell, A. Ganesh, A. McCallum, 'Energy and Policy Considerations for Deep Learning in NLP' (Cornell University 2019) < https://arxiv.org/abs/1906.02243> accessed 20/12/2021
[295] Ibid
[296] A. Lacoste, A. Luccioni, V. Schmidt, T. Dandres, 'Quantifying the Carbon Emissions of Machine Learning' (arxiv.org, 2019) < https://arxiv.org/abs/1910.09700> accessed 20/12/2021
[297] Ibid n244 p. 50
[298] Ibid n244 p. 50

aims and combat numerous technical regulatory issues (most of these regulatory are discussed within Chapter Five of this thesis).

Whilst this document sets out a number of potentially promising regulatory intentions, it remains primarily suggestive and visionary; it lays down very few concrete actions, especially when compared to the proposals of other nations. The US[299] and China[300] for example have all set out very clear regulatory plans, with some publishing considerable detail as to their intended regulatory mechanisms and schemes. The UK on the other hand lists a number of rather vague goals within its Strategy including creating a governance framework that is flexible but does not create unnecessary burdens, working with global partners to promote international agreements, and enabling AI products and services to be trustworthy.[301]

These are all very reasonable and worthy endeavours, although in parts it would be beneficial to see specifically how some of these plans will be achieved. The UK approach within this pillar appears to be rather similar to that taken by a number of other nations including Canada; Canada was actually the first country to publish a national AI strategy in 2017 which outlined plans similar to that of the UK. [302] As of yet the Canadian government have not implemented any strict regulations in this space (similar to those proposed within the EU), but their AI strategy does seem to have had some success with research, talent acquisition and AI start-up's benefitting.[303] Therefore, despite the rather broad approach taken by the UK within their Strategy, they may still see success albeit there may not be much immediate regulatory action taken. *Standards, international collaboration, and sector specific approach*

It is worth noting that headway in this space is also beginning to be made with the announcement of the new UK AI Standards Hub pilot.[304] The creation of this hub is hoped to enable the collaborative creation of technical specifications, codes of practice and internationally agreed and endorsed standards related to the creation and use of AI. The creation of this hub is promising and does signify the leading role the UK government wish to play in this space.

[299] J. F. Weaver, 'Regulation of Artificial Intelligence in the United States' in W. Barfield, U. Pagallo (eds) *Research Handbook on the Law of Artficial Intelligence* (Elgar, 2018)
[300] J. Schuett, 'Defining the scope of AI regulations' (2021) *LPP Working Paper Series No 9-2021*
[301] Ibid n244 p. 50
[302] CIFAR, 'Pan-Canadian National AI Strategy' (cifar.org) < https://cifar.ca/ai/> accessed 20/01/2022
[303] Ibid
[304] Department for Digital, Culture, Media and Sport, Office for Artificial Intelligence, 'New UK initiative to shape global standards for Artificial Intelligence' (GOV.uk 2022)

Acknowledgment by the UK government that standards should be integrated into digital regulation is key; by firstly playing a leading role in the development of these technical standards, any resulting legislation will be subsequently more globally applicable, recognisable, and interoperable.[305]

Interestingly, the Strategy notes that the UK intends to continue to explore its view that a sector-led approach is most applicable at this stage in the evolution of AI, as opposed to a blanket-style regulation.[306] Although, this position is likely going to be put to the test during the coming years, and its viability tested. This would mean that regulators within various sectors would be responsible for the governance of AI on a case-by-case basis, however, whether regulators and other bodies across sectors are equipped to take on this task is another question entirely. This increase in responsibilities for these sector-specific regulators may in fact drive need for further investment, increased powers and even internal reform in order to function effectively in this capacity.

3.2.2 How does the strategy embed the principles of accountability, transparency, and non-discrimination?

With accountability, transparency and non-discrimination being the most commonly agreed upon ethical AI principles, it is important to consider just how well the UK Strategy addresses these factors within its pillars. The UK Strategy considered here is not a legally binding measure by any means, it is primarily a signalling document and gives indication of the UK's intentions in this space in both the short and long term. One might expect to see more solid evidence of these ethical principles are going to be embedded in a proposal for legislation (such as the EU AI Act for example, see the following Chapter for further detail), in comparison to a strategy document.

However, there are a number of parts of this document that raise questions regarding the treatment of these principles. For example, Pillar One of the Strategy makes reference to data, particularly how the UK government are planning to improve and enable data sharing and "to permit the collection and processing of sensitive and protected characteristics data" in order to attempt to monitor the bias and discrimination problem present within AI.[307] This is positive as it provides that there is awareness within the UK government of the bias and discrimination problem caused by AI, and therefore there are some attempts within the Strategy to address the ethical principle of non-discrimination By association, this also

[305] Ibid n244 p. 56
[306] House of Lords Select Committee on Artificial Intelligence, 'AI in the UK: ready, willing and able?' (2018) HL Paper 100
[307] Ibid n244 p. 31

means that transparency and accountability are considered to an extent by the Strategy as the three principles are intrinsically linked; by improving transparency within AI systems, we are encouraging accountability and minimising risk of discrimination.

Despite this affirmation of the key ethical principles, it would appear as though the UK government are intending to move away from the GDPR-style framework to something slightly more flexible and generous.[308] GDPR is renowned for being one of the most rigorous and protective pieces of data protection legislation currently in force, and so a move away from this framework to a more lenient data protection legislation may in fact have a negative impact upon the promotion of the three key ethical AI principles. This is due to the likelihood that by introducing data protection measures that are less strict than GDPR and therefore more 'business-friendly', the UK may in fact encourage the development of less ethical AI; AI that isn't transparent regarding data collection and use, does not have clear accountability, and has the potential to be discriminatory.

This is only reinforced by the governments March 2023 White Paper which drives home their 'pro-innovation' approach to AI regulation.[309] This paper does acknowledge certain principles such as transparency and explainability as important factors that should be embedded in the AI development process, however, the paper does go on to say that both of these principles are in fact not absolute requirements as they are difficult to achieve, and should therefore be applied proportionately.[310]

Therefore, the answer remains unclear as to how well the current UK stance on AI regulation will incorporate accountability, transparency, and non-discrimination into future regulatory requirements on AI development and deployment. It is clear that these principles are on the radar of the UK government, and that they are acknowledged as important, however, whether or not they will be at the forefront of any UK legislative agenda remains to be seen.

3.2.3 Conclusion

The approach taken by the UK in their recently published National AI Strategy can therefore be seen as an ambitious one, containing a number of rather high-level goals with a few specific actions in mind to achieve them. As pointed out here, there are a number of noteworthy points to be gained from reading this document, primarily that it is a signalling document as opposed to a clear-cut call to arms. This means that whilst the document sets

[308] Ibid n244 p. 31

[309] Department for Science, Innovation and Technology, Office for Artificial Intelligence, 'A pro-innovation approach to AI regulation' (gov.uk, 2023) < https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper> accessed 05/04/2023

[310] Ibid

out the intentions for the government's handling of AI, both short and long-term, the Strategy lacks specific detail in parts as to how the overall aim of becoming an AI world leader might be achieved.

Although, as with other similar National Strategies employed by the UK and other countries also, it is possible that given ample time, the AI strategy proposed within the UK may in fact lead to meaningful regulation, or at least begin to establish its foundations. Therefore, it would appear as though the groundwork for a far-reaching AI regulatory model has been laid in the UK.

3.3 Regulatory Strategies in the United States (US)

Just a couple of days prior to President Biden taking office, the Trump administration introduced the National Artificial Intelligence Initiative Office which is to facilitate research and policymaking collaboration across sectors.[311] The creation of this Office was a requirement of the American AI Initiative (Executive Order 13859) introduced in February 2019, again, its purpose being to solidify the US as a world leader in the regulation of AI.[312] Unlike the proposed approach taken by the EU, there is no singular federal regulation for AI within the US. As such, despite the seemingly proactive approach taken by the US government to regulating AI signalled by Executive Order 13859, the US has generated very little guidance on their intentions to introduce blanket or federal regulation in this space in recent years.[313]

Interestingly, the proposed US strategy is quite similar to that suggested within the UK's National AI Strategy, one that deals with the regulation of AI on a sector-by-sector basis.[314] This was a concept first considered during the Obama administration, during which three reports were published on the issue of AI governance. Due to the law-making sovereignty of US states, this sector-by-sector, piecemeal approach in some ways suits the constitutional structure within the US. Thus far, the regulation of AI within the US can be broken down into three categories; firstly, initiatives being taken at federal level by different agencies targeting specific sectors, secondly, Bills enacted by specific states, and thirdly, regulations introduced that target specific AI technologies. Each of these categories of regulation will be explored

---

[311] Trump Whitehouse Archives, 'Artificial Intelligence for the American People' (trumpwhitehouse.archives.gov, 2022) < https://trumpwhitehouse.archives.gov/ai/executive-order-ai/> accessed 01/04/2022

[312] E.O. 13859 of Feb 11, 2019 84 FR 3967

[313] Y. Chae, 'U.S. AI Regulation Guide: Legislative Overview and Practical Considerations' (2020) *The Journal of Robotics, Artificial Intelligence & Law* 3(1) 17-40, 17

[314] Executive Office of the President, National Science and Technology Council, Committee on Technology, 'Preparing for the future of artificial intelligence' (obamawhitehouse.archives.gov, 2016) <https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf> accessed 01/04/2022

within this section, with the aim to evaluate the potential effectiveness of these differing approaches.

3.3.1 Department and Agency-lead AI initiatives

There are 15 departments that make up the executive branch of the US federal government, and over 400 agencies, most of which typically have legislative functions. First and foremost, in the list of departments engaging in work to regulate AI is the Department of Commerce, who are at present the most proactive development in the development of AI regulation within the US. During September 2021, the department announced the creation of a National Artificial Intelligence Advisory Committee (NAIAC) (launched in April 2022) which intends to advise on:

- U.S. AI competitiveness; progress in implementing the Initiative
- The state of science around AI
- Issues related to AI workforce, including barriers to employment supporting opportunities for historically underrepresented populations
- How to leverage initiative resources
- The need to update the initiative
- The balance of activities and funding across the initiative
- The adequacy of the National AI R&D Strategic Plan
- Management, coordination, and activities of the initiative
- Adequacy of addressing societal issues
- Opportunities for international cooperation
- Issues related to accountability and legal rights
- How AI can enhance opportunities for diverse geographic regions[315]

We are yet to see any outcomes from the NAIAC as its inaugural meeting was on 4th May 2022. However, coupled with the relationship that exists between the Department of Commerce and the National Institute of Standards and Technology (NIST), it seems as though NAIAC has the potential to influence considerably the shape of AI regulation within the US, specifically with regards to how businesses approach AI development and use. This is perhaps best evidenced by the AI Risk Management Framework currently being

[315] US Department of Commerce, 'Department of Commerce Establishes National Artificial Intelligence Advisory Committee' (commerce.gov, 2021) < https://www.commerce.gov/news/press-releases/2021/09/department-commerce-establishes-national-artificial-intelligence> accessed 01/04/2022

developed by NIST which aims to inform organisations on AI risks that should and can be avoided.[316]

The Risk Management Framework proposed by NIST was open for comment during Spring/Summer 2022, but the initial draft displayed its intentions clearly. Interestingly, this risk-based approach is in some ways similar to that proposed within the EU AI Act (which will be considered in detail within Chapter Four), in that it acknowledges risk as a potential factor to use in the regulation of AI, yet it approaches risk in a slightly different way. The NIST framework first identifies the general public as a stakeholder group and part of the audience for this governance strategy. It also frames risk in a slightly different way, by considering more closely the potential harms that might result from AI use, and then categorising these harms, e.g., harm to people, harm to organisations and harm to systems.[317] Therefore, this approach appears to more directly tackle the issues of accountability, transparency, and non-discrimination. Similarly, the framework considers the various characteristics of AI and the risks that accompany these, e.g., technical, socio-technical, and other guiding principles.[318]

Therefore, from the outset the NIST framework presents an innovative approach to AI risk and displays a promising approach to AI governance. Due to the nature of NIST as a standards development organisation (SDO), the finished product has the potential to be adopted by organisations across the globe as a foundational standard in AI use and development.

However, as discussed in Chapter Five of this thesis, there are draw backs to basing AI governance strategies entirely on industry standards, the most obvious shortcoming being that these standards are not usually legally binding. Therefore, whilst the proposed NIST framework is most definitely a positive step forward in the regulation of AI, to be entirely successful it is likely that there will need to be additional more prescriptive regulatory measures enforced in accompaniment.

In addition to the Department of Commerce, the Federal Trade Commission (FTC) appears to be poised to act in the regulation of AI.[319] In particular, the FTC are aiming to crack down

[316] NIST, 'AI Risk Management Framework: Initial Draft' (nist.gov, 2022) < https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf> accessed 01/05/2022
[317] Ibid
[318] Ibid
[319] H. Sussman, R. McKenney, A. Wolfington, 'U.S. Artificial Intelligence Regulation Takes Shape' (Orrick LLP, 2021) <https://www.orrick.com/en/Insights/2021/11/US-Artificial-Intelligence-Regulation-Takes-Shape#:~:text=Artificial%20Intelligence%20(AI)%20has%20the,next%20era%20of%20technological%20advancement.> Accessed 01/04/2022

on the use of biased algorithms, as made clear in their business blog post on algorithmic truth, fairness, and equality published in 2021.[320] Again, this post is aimed at businesses and their use of AI, and specifically highlights a number of FTC regulations that will be utilised to stop the use of unfair and biased decision-making algorithms, this includes section 5 of the of Federal Trade Commission Act, the Fair Credit Reporting Act, and the Equal Credit Opportunity Act.[321]

It is worth noting though that these regulations made use of by the FTC would appear to be retroactive in their approach to handling algorithmic bias as opposed to preventative. For example, the Fair Credit Reporting Act is intended to be put to use when an algorithm denies an individual employment, housing, or other benefits.[322] Similarly, the Equal Credit Opportunity Act functions in a similar way, making it illegal for an organisation to use an algorithm that discriminates based upon a protected characteristic.[323]

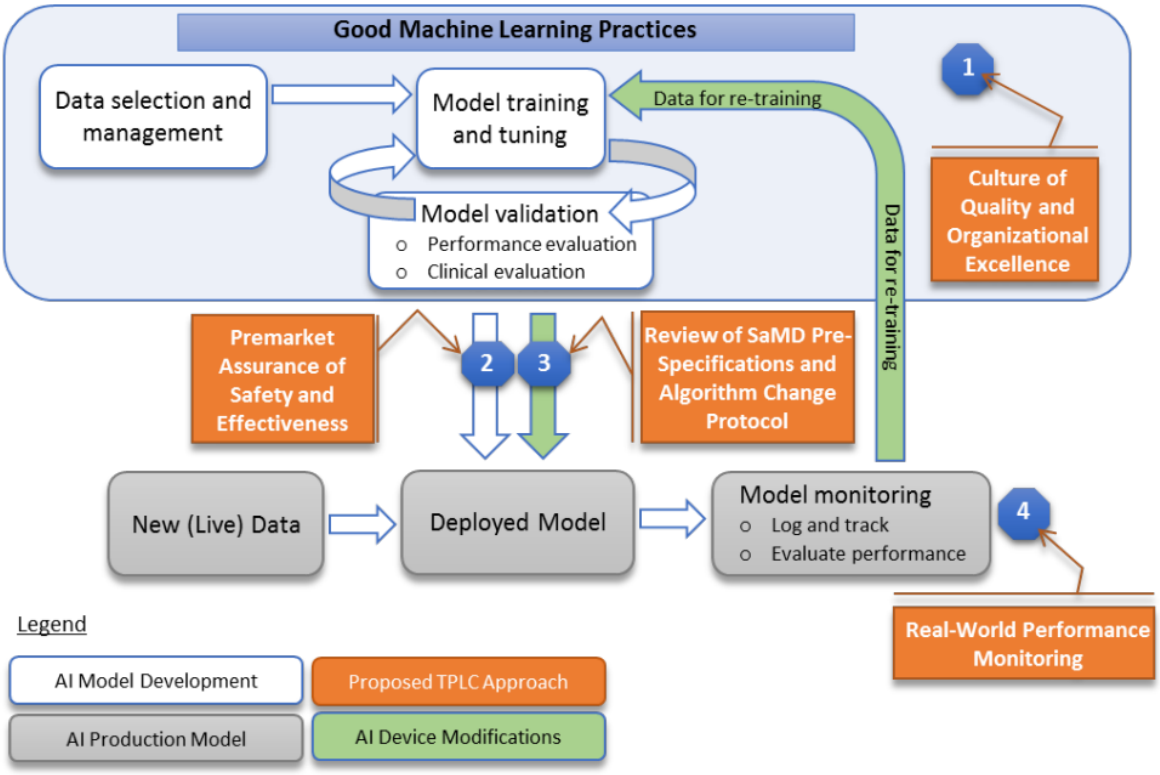
Therefore, whilst these legal provisions criminalise the use of algorithms that result in unfair and discriminatory outcomes, it does not entirely prevent harm from being caused; an algorithm of this sort still has to be developed, deployed and tested by the general public first of all to establish that it is biased and unfair in some capacity, thus causing harm before it will be deemed unsuitable for use and prohibited by the law. As a result, despite the apparent effort shown by the FTC to regulate this space to some degree, more work must be done to reduce AI-induced harm from the start, e.g., by developing more preventative measures as opposed to relying on retroactive ones such as this.

Further to this, associates at Orrick highlight another further US government agency working towards regulating AI, the Food and Drug Administration (FDA).[324] The FDA published their AI in medical device action plan in 2021 which specifically discusses how the FDA are planning to govern the use of AI and machine learning software within medical devices, in particular software and devices used to diagnose illnesses, suggest treatment plans and mitigate disease.[325] As AI is relied upon more and more within the health care sector, whether it is for the reasons stated above or for use in everyday health tracking devices such as smart watches and health monitors, it is imperative that there is an oversight body

[320] E. Jillson, 'Aiming for truth, fairness, and equity in your company's use of AI' (ftc.gov, 2021) <https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai> accessed 01/04/2022
[321] Ibid n319
[322] Ibid n319
[323] Ibid n319
[324] Ibid n319
[325] U.S. Food and Drug Administration, 'Artificial Intelligence and Machine Learning in Software as a Medical Device' (fda.gov., 2021) <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> accessed 01/04/2022

equipped to ensure that any AI-based systems here are being used safely. One particular point of concern flagged by the FDA in their 2019 discussion paper on the topic, was how the agency could go about ensuring transparency within AI-based medical devices.[326]

Figure 3: FDA Proposal for Good Machine Learning Practices for Medical Devices[327]



Source: U.S. Food and Drug Administration, 'Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) Based Software as Medical Devices (SaMD), Discussion Paper and Request for Feedback' (2019) https://www.fda.gov/media/122535/download, accessed 01/04/2022

The above figure forms part of the FDA proposed framework on artificial intelligence-based software as medical devices. It demonstrates a system of consistent validation and retraining of an AI-based system using new and relevant data to ensure effectiveness. The system proposed here is promising, and could perhaps be adapted on a wider scale, for example within a blanket-style AI regulation. This could potentially work well as it takes into account the ever-changing nature of AI, and its ability to learn from its surrounding environment and data.

By exploiting this factor, the individuals tasked with deployment of the AI-based system have the opportunity to continually assess the systems effectiveness and finely tune the device to

[326] U.S. Food and Drug Administration, 'Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) Based Software as Medical Devices (SaMD), Discussion Paper and Request for Feedback' (fda.gov, 2019) <https://www.fda.gov/media/122535/download> accessed 01/04/2022
[327] Ibid p. 8

increase its functionality and ensure its reliability. Notably, this system does depend on a culture of quality and organisational excellence, as highlighted in the top right-hand corner of this figure.

The various publications, papers and proposals made by these government agencies and departments demonstrates an awareness that AI poses risks to our society that cannot and should not be ignored. There does seem to be a particular effort here to encourage companies to begin proactively thinking about how they currently use, and will continue to use, AI within their businesses. As discussed already, an agency-by-agency structured regulatory framework does appear to be quite piecemeal and so it is worth considering how these government department and agency-lead initiatives might work alongside both state-backed bills and other regulations targeting use of AI in specific sectors e.g., within the automotive industry.

3.3.2 State-by-state AI regulation

As per the constitutional structure of the US, each state retains its sovereignty to enact laws written by the states legislature and signed by the state Governor. Whilst there is no current federal AI regulation within the US, there are a plethora of state-led initiatives that have been introduced since 2019.

There were a total of thirty-three AI regulatory measures proposed by seventeen states in 2021, with only six of those successfully enacted within the year, twenty-one still pending, and six failed.[328] These proposed regulatory measures vary in nature, however they could all be categorised as general AI regulations (as opposed to ones that target a specific sector e.g. driverless cars etc.).[329] Those that were successfully enacted during 2021 tend to either establish a form of task force that will advise on the current state of AI-related harm (enacted within Alabama[330], Washington[331] and Illinois[332]), specifically prohibit the use of algorithms that may result in discrimination (enacted within Colorado[333] and Illinois[334]), or mandate that school curriculum includes teaching on subjects such as robotics, AI and machine learning (ML) as per Mississippi[335].

[328] National Conference of State Legislatures, 'Legislation Related to Artificial Intelligence' (ncsl.org, 2022) <https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx> accessed 15/05/2022
[329] Ibid
[330] AL 2021-344 SB78
[331] WA 2021 S.B. 5092
[332] IL 2021 H.B. 645
[333] CO 2021 S.B. 169
[334] IL 2021 H.B. 53
[335] MS 2021 H.B. 633

The Bills pending authorisation are quite varied. Hawaii for example is hoping to grant tax relief for investments in companies that are developing AI and cybersecurity products[336], whilst Massachusetts appears to be primarily concerned with data privacy and government transparency[337]. As such, the array of issues covered by current and proposed state regulations on AI again show an acknowledgement that this is a subject in need of governance but approaches to do so may vary. From this analysis conducted by the NCSL, it is also clear that some states appear to be less willing to enact regulatory measures for AI than others, for example measures introduced in both Missouri and Virginia failed in 2021.

It is proven that the political polarization that has existed within both Congress and at state legislative levels since the inception of the United States has proven to impact upon policies and legislative agendas.[338] It is interesting to therefore consider to what extent the politics of each state is impacting upon their approach to regulating AI, and ultimately the overall effect that politics is having upon the AI regulatory landscape in the US as compared with other nations. The answer, however, is not as clear cut as initially thought; in the example given above, it appears as though Missouri (often named a bellwether state) has voted Republican for the past two decades, whilst Virginia (previously a swing-state) has more recently become a predominantly blue-leaning or Democrat voting state.[339]

Moreover, there have been instances in which both Democrats and Republican lawmakers have unanimously agreed upon the need to govern AI, particularly facial recognition technology.[340] On the other hand, in preparation for the 2020 presidential election, both sets of candidates had rather little to say about AI other than acknowledging its need to be governed.[341] Therefore, it can be concluded that whilst it is proven that state politics do have an impact upon the success of a legislative agenda, it appears as though when it comes to AI both political parties are relatively balanced; both acknowledge the need to regulate but both reveal very little detail regarding their plans to do so.

---

[336] HI H.B. 454
[337] MA H.B. 136, MA H.B. 142, MA S. 60
[338] National Conference of State Legislatures, 'State Legislative Policymaking in an Age of Political Polarization' (ncsl.org) <https://www.ncsl.org/Portals/1/HTML_LargeReports/Partisanship_1.htm> accessed 15/05/2022
[339] D. F. Damore, R. E. Lang, K. A. Danielsen, *Blue Metro's, Red States: The Shifting Urban-Rural Divide in America's Swing States* (Brookings Institution Press, Washington D.C. 2020)
[340] D. Harwell, 'Both Democrats and Republicans blast facial-recognition technology in a rare bipartisan moment' (washintonpost.com, 2019) < https://www.washingtonpost.com/technology/2019/05/22/blasting-facial-recognition-technology-lawmakers-urge-regulation-before-it-gets-out-control/> accessed 15/05/2022
[341] G. Pooya, 'The great AI debate: What candidates are (finally) saying about artificial intelligence' (thehill.com, 2019) < https://thehill.com/opinion/technology/473794-the-great-ai-debate-what-candidates-are-finally-saying-about-artificial/> accessed 05/04/2022

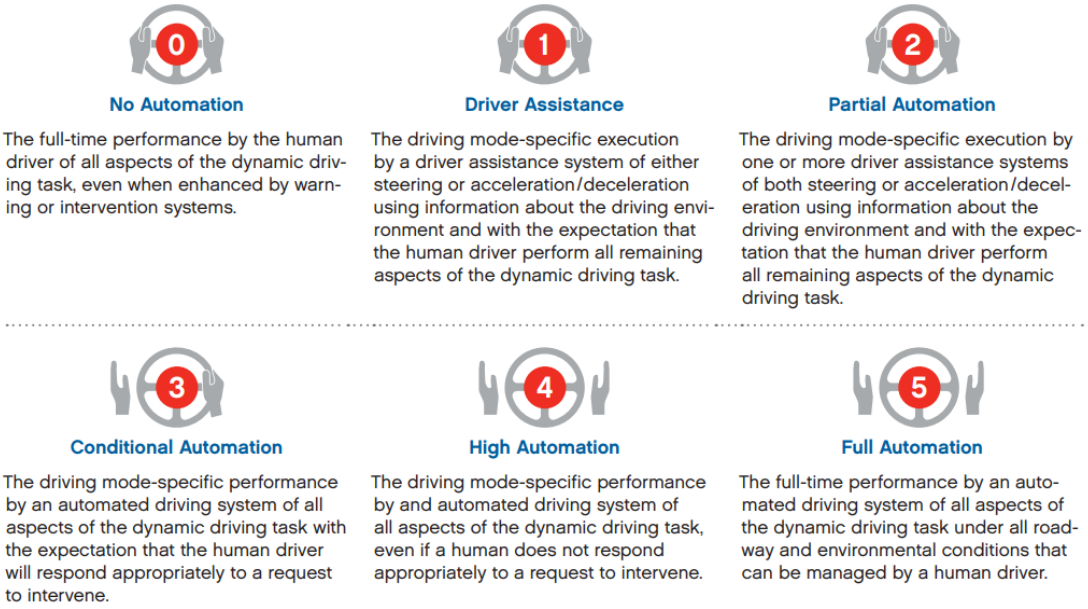
3.3.3 Sector-specific regulations

As discussed to an extent already, the final type of AI regulation categorised within this thesis is sector-specific regulation. This approach is favoured within the UK's National AI Strategy[342] and appears to be the favoured approach within the US. During both the most recent Obama administration and the Trump administration, particular focus was placed upon AI use within three sectors; education (promotion of STEM programs), military defence, and the autonomous vehicle industry.[343] The latter of these sectors, the autonomous vehicle industry, has arguably seen the most attention and regulatory action in recent years with a total of 76 Bills introduced (inclusive of those enacted, pending and failed).[344]

These Bills concern all manner of topics related to autonomous vehicle use, including the cybersecurity of the vehicle, vehicle infrastructure, licensing and registration, vehicle data, vehicle testing and vehicle operation.[345] The regulations in place also hinge upon the level of automation in question, as demonstrated via Figure 6. This typically means that vehicles with a higher degree of automation will require a higher level of governance and oversight.

[342] Office for AI, Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, 'National AI Strategy' (GOV.uk, 2021) <https://www.gov.uk/government/publications/national-ai-strategy>accessed 05/04/2022
[343] R. Girasa, 'AI U.S. Policies and Regulations' in *Artificial Intelligence as a Disruptive Technology* (Palgrave Macmillan, London 2020) 69-102
[344] National Conference of State Legislatures, 'Autonomous Vehicle State Bill Tracking Database' (ncsl.org, 2022) < https://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx> accessed 20/07/2022
[345] Ibid

Figure 4: Society of Automotive Engineers Automation Levels[346]



Source: Jones Day, 'White Paper, Autonomous Vehicles: Legal and Regulatory Developments in the United States' (2021) https://www.jonesday.com/-/media/files/publications/2021/05/autonomous-vehicles-legal-and-regulatory-developments-in-the-us/files/autonomous-vehicles-legal-and-regulatory-developme/fileattachment/autonomous-vehicles-legal-and-regulatory-developm.pdf, accessed 20/07/2022

Interestingly though, in March 2022 the National Highway Traffic Safety Administration (NHTSA) significantly revised current regulations regarding the expected makeup of a driverless vehicle following a petition from General Motors.[347] In doing so, the NHTSA removed regulations that require an autonomous vehicle to always have a driver's seat, steering wheel and other similar features as they were deemed to be 'logically unnecessary'.[348] In essence, this removes from current autonomous vehicle regulations the requirement for there to be any human driver capable of taking control of the vehicle from inside whilst in tandem.

This seems to be somewhat of a transgression in the promotion of autonomous vehicle safety, but progression in encouraging the use of this type of tech in everyday life. In terms of autonomous vehicle safety, it would be worthwhile considering whether it is the intention of the NHTSA is to introduce more stringent testing regulations prior to deployment of the vehicles in the absence of the recently removed rules, however at present it is unclear as to

[346] Jones Day, 'White Paper, Autonomous Vehicles: Legal and Regulatory Developments in the United States' (jonesday.com, 2021) < https://www.jonesday.com/-/media/files/publications/2021/05/autonomous-vehicles-legal-and-regulatory-developments-in-the-us/files/autonomous-vehicles-legal-and-regulatory-developme/fileattachment/autonomous-vehicles-legal-and-regulatory-developm.pdf> accessed 20/07/2022

[347] Department for Transport, National Highway Traffic Safety Commission, 49 CFR Part 571, Docket No. NHTSA-2021-0003, RIN 2127-AM06 Occupant Protection for Vehicles with Automated Driving Systems

[348] Ibid

whether they will introduce such measures. There should hopefully be some comment on this in the near future, seeing as during 2021 it was reported that there were 9.1 autonomous vehicle crashes per million miles driven as compared with 4.1 human-driven vehicle crashes, making the rate at which autonomous vehicles crash over double that of a regular vehicle.[349]

It would also be worthwhile considering public opinion surrounding the use of autonomous vehicles, given the removal of these rules; this is due to the fact that most government-led AI strategies seem to place some priority in improving the trustworthiness of AI. A study conducted in 2021 on public perceptions of autonomous vehicle use show some interesting findings, and despite taking place prior to the recent NHTSA ruling, they show some useful data. 84% of respondents said that they would still prefer to drive a car as opposed to riding in one driven autonomously, and 43% of respondents believe that there should be a set of nationally consistent regulations created by the Department of Transportation in comparison to 18% who would prefer state regulations.[350]

One potential reason for the higher percentage of support for federal action in this space is based upon the geography of the US. Crossing state borders in the US is a relatively easy task, particularly in areas such as the Northeast, in such a case it would be logical to have a consistent set of national rules and regulations to govern the requirements for autonomous vehicles crossing borders.

Most interestingly, 76% of respondents said that they either favour or strongly favour a person being required in the driver's seat of the vehicle who could take control if necessary.[351] This thinking is in stark contrast with the current NHTSA revisions and will most likely decrease public trust in the use of AI in this sense, with over half the respondents claiming development of this type of tech makes them feel worried.[352] Perhaps a way to alleviate this mistrust might be to make it abundantly clear that other equally robust provisions will be put in place to ensure the safety of those using these vehicles e.g. via enforcing more stringent testing regulations (further consideration on this point can be found within Chapter Five of this thesis). Whilst this specific example might not be directly related to issues of human rights violations as a result of AI use, it still demonstrates the lack of trust we currently see in AI-based systems.

---

[349] Clifford Law Offices, 'Driverless Car Accidents – Who is at Fault?' (2021) *National Law Review* 11(176) https://www.natlawreview.com/article/driverless-car-accidents-who-s-fault accessed 20/07/2022

[350] E. Taylor, 'Autonomous Vehicle Decision-Making Algorithms and Data-Driven Mobilities in Networked Transport Systems', *Contemporary Readings in Law and Social Justice* 13(1) 9

[351] Ibid

[352] Ibid

3.3.4 How do these approaches embed the principles of accountability, transparency, and non-discrimination?

There are some key takeaways from the US approach to AI regulation with regards to how they incorporate ethical principles such as accountability, transparency, and non-discrimination. For example, the NIST Risk Management Framework discussed at the beginning of this section frames risk in a unique way, by considering more closely the potential harms that might result from AI use, and then categorising these harms, e.g., harm to people, harm to organisations and harm to systems.[353]

By identifying and categorising harm in this way, the framework is more effectively addressing ethical principles; e.g. if via this framework an AI-based system can be identified as having the potential to infringe upon a person's privacy due to improper data use, it can be interpreted that there is likely a lack of transparency in the system, we can now establish accountability and minimise risk of discrimination. Therefore, it would seem that frameworks like this one may go some way in helping us to ensure key ethical principles such as those mentioned here are upheld.

However, there is still room for improvement. It has been made clear by US government departments such as the FTC that they intend to use existing law to combat harmful AI use. The issue with this is that many of the measures that are deployed by the FTC would be retroactive. This means that whilst these measures, such as the Fair Credit Reporting Act 1970, might help to establish accountability for harm caused by an AI and as such punish an organisation for deploying a discriminatory algorithm, it will not necessarily prevent the harm from happening. Therefore, whilst there is effort on behalf of the US to promote ethical AI principles like accountability, transparency, and non-discrimination, there is still room for improvement. Perhaps more explicit reference within AI regulatory efforts to instruments such as the American Convention on Human Rights would be beneficial,[354] and how regulatory efforts strive to uphold the rights contained in such documents would be valuable.

3.3.5 Conclusions

The US has taken a rather unique approach so far to the regulation of AI. It could be most closely compared to the approach proposed by the UK, in that they favour a sector-led approach to AI governance. Although, it is fair to say that AI regulations within the US are slightly more progressed than any within the UK.

---

[353] Ibid n319

[354] Inter-American Commission on Human Rights, American Convention on Human Rights, Adopted at the Inter-American Specialized Conference on Human Rights, San José, Costa Rica, 22 November 1969

The sector-led approach does have its merits; the regulations will likely be more tailored-made to specific uses and applications for example medical use or for the manufacture of autonomous vehicles, and it is more likely that the use of AI within a given sector will be more closely monitored by an oversight body, e.g., the NHTSA. As a result, it seems as though accountability is a key focus of the US regulatory approach to AI, particularly considering the example of automated vehicles. Although, it does have its downfalls; it makes for a more fragmented approach to AI regulation which may lead to uncertainty, and for certain AI applications this may present difficulties in cross-border use (e.g., autonomous vehicles).

All in all, we have a relatively clear picture as to what the UK and the US are doing at present in terms of AI regulation, all of which seem to place it relatively at the forefront of their legislative agendas. However, it is worthwhile considering what other nations are doing in this field, namely China.

3.4 Regulatory Strategies in China

In a similar vein to the UK and US, China issued its New Generation Artificial Intelligence Development Plan (NGAIDP) in 2017 to be guided by its AI Strategy Advisory Committee.[355] The plan highlights China's aim to become world leading in AI development within the next decade or so, a similar goal to that of the other nations previously discussed in this Chapter. Although, the 2017 development plan is not the first policy document to be released by the Chinese government on the point of AI.

In 2015 the State Council released guidelines that aimed to promote research and development in this field, the 2015 'ten-year plan' also included the aim for China to become world leading in AI manufacturing, and the 2016 'five-year plan' declared AI as critical for stimulating economic growth.[356] Prior to 2015, AI was not necessarily a topic included within any of China's legislative agendas. However, with the NGAIDP being the most recent and most prescriptive policy document released by China on their current intentions in this space, it is worthwhile considering it in further detail.

[355] State Council, 'Notice of the State Council on Issuing the New Generation Artificial Intelligence Development Plan' (gov.cn, 2017) <http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm> accessed 10/05/2022

[356] H. Roberts, J. Cowls, J. Morley, M. Taddeo, V. Wang, L. Floridi, 'The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation' (2021) *AI & Society* 36, 59-77

3.4.1 New Generation Artificial Intelligence Development Plan 2017

The plan sets out to achieve its aims by 2030, the main plan being to place AI as the 'driving force' for China's economic transformation.[357] This goal has already been achieved to some extent with the Chinese industrial sector generating around 32.6% of the country's GDP in 2021, making it the largest contributing sector when compared to retail or finance for example.[358] Not only this but the plan highlights the invaluable role that AI can and will play within society, as it projects use within welfare settings, environmental protection, medical care and judicial services within the next decade.[359] The use of AI in these settings within China is not a new phenomenon, for example the Chinese Social Credit System that has been in development since the early 2000's utilises AI in such a way that overlaps many of these varying parts of society.[360]

The plan sets out three key dates by which particular milestones must be reached, which are as follows:[361]

1. By 2020, China will have achieved important progress in the development of the AI industry. AI will have become important in the country's economic growth, and "AI technology applications will have become a new way to improve people's livelihoods".
2. By 2025, will establish itself as a true world leader in AI by achieving breakthroughs in AI theory and development. Within this part of the plan, China plan to have implemented AI laws, regulations, and ethical norms, as well as technical standards.
3. By 2030, China will be the 'world's primary AI innovation centre'. They also mention at this point that they will have achieved breakthroughs in 'brain-inspired intelligence', what this means is yet to be discovered.

As pointed out by Roberts et al[362], the plan is not actually intended to function as a centralised initiative, but rather as a means for encouraging technological innovation amongst the private sector and local governments. Sheehan therefore labels the plan as more of a 'wish-list' as opposed to an actual AI strategy adopted by other nations and

[357] Ibid
[358] Statista, 'Distribution of the gross domestic product (GDP) in China in 2021' (statista.com, 2022) < https://www.statista.com/statistics/1124008/china-composition-of-gdp-by-industry/#:~:text=In%202021%2C%20the%20industrial%20sector,of%20the%20country's%20economic%20output.> accessed 10/05/2022
[359] Ibid n355
[360] D. Mac Síthigh, M. Siems, 'The Chinese Social Credit System: A model for other countries?' (2019) *Modern Law Review* 82(6)
[361] Ibid
[362] Ibid n356

agencies.[363] This is furthered by the fact that as a part of this plan specific businesses were selected as 'AI national champions', these companies would focus on specific government strategies (as included within the NGAIDP) and as a result would receive preferential treatments with regards to governments bids and funding.[364]

At the time of writing, we have bypassed the first key date listed within the NGAIDP and are well on our way to the second date listed, so we should now be able to see whether the first key objectives set out within the plan been achieved. In 2019 there was a fall in investment in AI which was only perpetuated by the following COVID-19 pandemic, this has slowed progression down slightly however this does not mean that many of the goals set to be achieved by 2020 have not been met.[365] As pointed out by Zeng[366], there has been significant progression in provincial governments across China continuing to adopt 'intelligent government' models; this includes utilising facial recognition technology to streamline and improve a number of services, yet the privacy ramifications of this are very scarcely considered.

3.4.2 Other AI initiatives

Interestingly, during the latter part of 2021 China have released a number of documents that build upon their initial NGAIDP. These initiatives are interesting as they predominantly deal with the regulation of AI, as opposed to purely encouraging economic growth in the sector. It seems likely that this is an attempt to meet in part the key strategic objectives set out within the NGAIDP to be reached by 2025.[367] These three initiatives focus on:

1. Rules for online algorithms[368]
2. Testing and certification of AI systems, with a focus on promoting trustworthiness[369]

[363] M. Sheehan, 'How China's Massive AI Plan Actually Works' (marcopolo.org, 2018) <https://macropolo.org/analysis/how-chinas-massive-ai-plan-actually-works/> accessed 10/05/2022
[364] A. Graceffo, 'China's National Champions: State Support Make Chinese Companies Dominant' (foreignpolicyjournal.com, 2017) <https://www.foreignpolicyjournal.com/2017/05/15/chinas-national-champions-state-support-makes-chinese-companies-dominant/> accessed 15/05/2022
[365] B. Horton, J. Zeng, 'Can China become the AI superpower?' (chathamhouse.org, 2021) <https://www.chathamhouse.org/2021/01/can-china-become-ai-superpower> accessed 15/05/2022
[366] Ibid
[367] M. Sheehan, 'China's New AI Governance Initiatives Shouldn't Be Ignored' (carnegieendowment.org, 2022) < https://carnegieendowment.org/2022/01/04/china-s-new-ai-governance-initiatives-shouldn-t-be-ignored-pub-86127> accessed 15/05/2022
[368] Cyberspace Administration of China, 'Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms' (digichina.stanford.edu, 2021) No. 7 < https://digichina.stanford.edu/work/translation-guiding-opinions-on-strengthening-overall-governance-of-internet-information-service-algorithms/> accessed 10/05/2022
[369] Center for Security and Emerging Technology, 'White Paper on Trustworthy Artificial Intelligence' (cset.georgetown.edu, 2021) < https://cset.georgetown.edu/publication/white-paper-on-trustworthy-artificial-intelligence/> accessed 10/05/2022

3. Establishing AI ethical principles[370]

It is promising to see development in the regulation of AI as opposed to policy purely focusing on industrial growth, which is what we have seen this far. It would appear that out of the three initiatives, the first is the most well-developed and has already produced significant results; for example, in late 2021 the Cyberspace Administration of China (CAC) produced a set of rules for governing internet recommendation algorithms.[371] The provisions here are quite prescriptive, requiring service providers to include interfaces that allow autonomous choice alongside their algorithms, provide transparency to users, and offer an option for users to switch of algorithmic recommendation systems, to name a few.[372]

Some of these provisions, such as those listed above that require service providers to provide some kind of explanation behind the content being shown to them, are very progressive and could actually provide somewhat of a blueprint to be followed by other nations and regulators, as in turn this will likely help to improve AI trustworthiness which seems to be goal for most involved in this space. Markedly, the three regulatory initiatives announced by these Chinese administrations are likely to be in competition with one another. Often these bodies are competitive, each striving to achieve more than the other in order for their policies to be endorsed centrally.[373] At present it would appear that CAC are leading this race progressing even further than the UK and US have in the debate on algorithmic transparency so far.

The aforementioned CAC recommendations for internet algorithm management came into effect on 1st March 2022, making it the first real global effort to regulate the use of AI in this way. The regulations will effectively govern the use of algorithms to recommend specific content online, and similarly will prevent companies from offering services and products for different prices based upon a user's personal information.[374] This means that companies breaching these rules can be fined or even have their websites and online services shut down and banned entirely in China.[375] The regulations also look to be relatively successful,

---

[370] Center for Security and Emerging Technology, 'Ethical norms for New Generation Artificial Intelligence Released' (cset.georgetown.edu, 2021) < https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/> accessed 10/05/2022
[371] Cyberspace Administration of China, 'Internet Information Service Algorithmic Recommendation Management Provisions (digichina.stanford.edu, 2022) < https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/> accessed 10/05/2022
[372] Ibid
[373] S. V. Lawrence, M. F. Martin, 'Understanding China's Political System', *Congressional Research Service* (sgp.fas.org, 2013) < https://sgp.fas.org/crs/row/R41007.pdf> accessed 10/05/2022
[374] J. Conrad, W. Knight, 'China is About to Regulate AI – and the World is Watching' (wired.com, 2022) < https://www.wired.com/story/china-regulate-ai-world-watching/> accessed 10/05/2022
[375] Ibid

even prior to their formal implementation in March 2022, with a number of popular Chinese apps introducing ways for users to opt out of algorithmic recommendations.[376]

3.4.3 How do these initiatives embed the principles of accountability, transparency, and non-discrimination, and is this feasible in the case of China?

It is clear therefore that the principles of transparency and accountability are being addressed to some extent within recent Chinese AI initiatives, with the CAC rules on internet recommendation algorithms providing for increased transparency for consumers, and clear accountability for organisations found to be violating the rules. Yet over the past decade China has been the subject of much debate regarding AI use by its authoritarian regime, which has been able to increase state control via digital means.[377] As a result, it remains questionable as to how feasible it actually is that key ethical principles such as accountability, transparency and non-discrimination would be fully embedded within Chinese AI.

As discussed previously in this section, China has been developing and using (or misusing) AI for a long time, China's social credit system is a prime example of this; a means to digitally control a nation.[378] Whilst AI is being leveraged by the Chinese Communist Party to assert better state control, there is also evidence that the government used facial recognition technology to track down and control the Uighur community, and move them into detention camps.[379] This is the first case of government intentionally using AI-based technology to carry out racial profiling and discriminate against an entire community, which is in obvious contention with all ethical principles associated with AI.

However, as put by Zhu, despite China having far fewer domestic AI ethics guidelines than other nations and regions around the world, there are diverse discussions taking place regarding AI ethics in China.[380] It is therefore proposed that rather than questioning whether these discussions are taking place, we should focus on their content and substance.[381] It is also worth noting that the majority of the conversation taking place regarding AI ethics in

---

[376] Baidu, 'App cannot force personalised recommendation, do you know how to close it?' (baijiahao.baidu.com, 2021) <https://baijiahao.baidu.com/s?id=1709066811285781912&wfr=spider&for=pc> accessed 10/05/2022
[377] J. Zeng, 'Artificial Intelligence and China's authoritarian governance' (2020) *International Affairs* 96(6) 1441-1459 *1442*
[378] Ibid n358
[379] P. Mozur, 'One Month, 500,000 Face Scans: How China Is Using AI to Profile a Minority' (nytimes.com, 2019) < https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html> accessed 05/04/2023
[380] J. Zhu, 'AI ethics with Chinese characteristics? Concerns and preferred solutions in Chinese academia' (2022) *AI & Society* < https://link.springer.com/article/10.1007/s00146-022-01578-w> accessed 05/04/2023
[381] Ibid

China is happening within academia, as the majority of those involved in ethical committees and engaging in international exchanges on AI regulation are Chinese scholars and academics.[382] This work being undertaken is done in alignment with the government initiatives discussed earlier in this section, primarily to establish China as an AI innovator, and so there is some to degree to which the Chinese government are involved in these ethics discussions.

Therefore, we are left with somewhat of a contradiction regarding how feasible it is for ethical principles such as accountability, transparency, and non-discrimination to be embedded within Chinese AI governance. There are rules such as those imposed by the CAC which increase transparency and accountability in AI and are therefore promising developments in terms of Chinese AI ethics. Yet simultaneously there are blatant and inexcusable examples of China misusing AI in order to discriminate and segregate an entire ethnic community. It would seem that China has the capacity and ability to truly invest in AI ethics and embed key ethical principles within their technology and regulation, however, whether these ethical interests outweigh those of the Chinese government will be the determining factor here.

3.5 Conclusions

It has been remarked that China is moving considerably faster than their counterparts in the UK and US.[383] However, this is not necessarily bad; in these nations, debates surrounding the correct way to regulate AI, proposed guidelines, rules and Acts have been circulating for a number of years without any meaningful and impactful regulations implemented. Whereas the aforementioned CAC recommendations for internet algorithm management were published, made available for comment, and enforced within a significantly short period of time.

It is arguable that this a reasonable way to approach AI regulation; when dealing with a consistently changing type of tech it is logical to also regulate at pace, as opposed to letting the tech change to such a high degree that the regulations initially proposed are no longer suitable. Therefore, the velocity with the Chinese government and its administrations are approaching the regulation of AI should perhaps serve as a lesson for other states wishing to regulate, that slower isn't always better.

3.5 Regulatory Strategies in South Africa

Building on the examination of UK, US and Chinese regimes, approaches to AI within South Africa offer a different and valuable perspective on AI governance. Therefore, this analysis is

[382] Ibid
[383] Ibid n374

rather different from the other analysis that has taken place so far within this Chapter, as there are no clear-cut legislative efforts to look at which specifically govern AI in South Africa. Yet, the South African perspective is crucial in developing our understanding of the differing interests that nations have in AI development, and what their priorities are in terms of governance.

As put by Ormond, South Africa has a significantly unequal society, this is evident via the vast digital divide prevalent there, and as such the country is on the "periphery of AI development, utilisation and regulation".[384] These factors therefore impact upon the interests that a nation has in regulation and sheds some light on the potential legal and ethical issues that AI might cause in that specific nation. This section considers recent efforts made in South Africa to deal with AI-related threats, and how South Africa intends on combatting related issues such as the digital divide.

3.5.1 Current State of AI Regulation in South Africa

At present, South Africa has no specific laws that regulate AI.[385] This does not mean that regulation is not necessary here or that it is not a topic for discussion; there are numerous AI related issues that are unique to the country and its neighbours for which some form of regulation is required, and solutions are being considered. Use of AI in South Africa is just as prevalent as in any other nation, which means there are also examples of AI producing concerning consequences in this nation.

One example of which was the deployment of an AI-driven digital ID system which was responsible for providing citizens with access to social grants.[386] These social grants are received by around 18 million South Africans, these people being some of the country's most vulnerable citizens. The company that created the digital ID system had access to the personal information of the recipients and used this data to send 'predatory financial offers' to them.[387] The company were then able to deduct loan repayments directly from the grants prior to them even reaching the beneficiaries, meaning some of these individuals were receiving little to no grant payments each month.[388]

---

[384] E. Ormond, 'Artificial intelligence In South Africa comes with special dilemmas – plus the usual risks' (2023) *The Conversation* <https://theconversation.com/artificial-intelligence-in-south-africa-comes-with-special-dilemmas-plus-the-usual-risks-194277> accessed 20/03/2023
[385] D. Donnelly, 'First Do No Harm: Legal Principles Regulating the Future of Artificial Intelligence in Health Care in South Africa' (2022) *Potchefstroom Electronic Law Journal (PELJ)* 25(1) 1-43 < https://dx.doi.org/10.17159/1727-3781/2022/v25i0a11118> accessed 20/03/2023
[386] R. McAdams, 'AI in Africa: Key Concerns and Policy Considerations for the Future of the Continent' (afripoli.org, 2022) < https://afripoli.org/ai-in-africa-key-concerns-and-policy-considerations-for-the-future-of-the-continent> accessed 01/04/2023
[387] Ibid
[388] Ibid n386

This case was the subject of a number of court cases, all of which highlighted the unethical sharing of data, the lack of preparedness of the South African government in dealing with instances such as this, and the lack of awareness that citizens have of their information rights.[389] Therefore, as with as with all other nations around the world, more needs to be done here to secure the rights and interests of the South African people from the negative impacts of AI.

However, the lack of regulation at present offers the South African government a valuable opportunity to tackle these socio-technological issues, most of which form the basis of the more common ethical issues we see stemming from AI, such as accountability, transparency, and discrimination. By essentially grappling with the root causes of these well-established ethical issues first, South Africa might well position themselves as a leading nation in AI regulation.

3.5.2 Presidential Commission on the Fourth Industrial Revolution

Therefore, it is necessary to look at available policy documents that signal the priorities of the South Africa with regards to artificial intelligence, and potential next steps the country might make in terms of regulatory measures. A good place to start is the document that sets out the government's approach to the Fourth Industrial Revolution, which is inclusive of AI and was published by the Department of Telecommunications and Postal Services in 2019.[390]

In a similar vein to the UK strategy discussed earlier in this Chapter, this document is a signalling piece that sets out the vague parameters of the South African governments intended role in this AI-related matters. As a result, the document is neither legally enforceable in nature nor detailed enough to give a fully comprehensive idea as to what a fully formed AI policy might look like. The document does contain a few references to some priorities of the Presidential Commission, including the advancement of research programmes that further knowledge of modern technologies, enhancing South Africa's global competitiveness in this space, and developing the skills of the public in order to enhance employability.[391]

These are logical priorities, and some are quite similar to those stated within the UK's strategy discussed earlier in this Chapter, particularly the furthering of research in the space

[389] Ibid n386
[390] Department of Telecommunications and Postal Services, 'Presidential Commission on the Fourth Industrial Revolution' (gov.za, 2019) <https://www.gov.za/sites/default/files/gcis_document/201812/42078gen764.pdf> accessed 04/04/2023
[391] Ibid n390

and working to develop tech related skills amongst the population. Albeit there is little reference to specifically how the South African government intends to tackle many prominent tech-related issues such as accountability, transparency, or discrimination. This isn't to say that there aren't calls for clearer statutory schemes that deal specifically with these issues, there certainly are there are requests for clearer guidance on liability and accountability resulting from AI use within health care.[392]

3.5.3 Dealing with socio-technological issues in South Africa

As identified earlier in this section, one of the primary findings of this analysis is that AI poses a number of specific threats within South Africa; there is a shortage of representative data to train AI, lack of AI-literacy and the potential for a deepened digital divide. These threats aren't necessarily as prevalent or as problematic in other nations, specifically nations situated in the global north.[393] Therefore, before any meaningful AI regulatory action can take place in South Africa, these issues need to be addressed fully. Doing so means that there is a real opportunity to lessen the digital divide that exists within South African society and allow the nation to fully benefit from new emerging technologies such as AI.[394]

There are a number of reasons why the digital divide is so prevalent in South Africa and other countries within the global south, but primarily it is due to the stark contrast in access to new technology in rural areas compared to urban areas and cities.[395] Although access to technology has generally improved over time, internet access and other technological infrastructures often come to rural communities much later than they do to urban areas, providing for unequal access to technology. It is also often the case that those within rural communities will have poorer educations than those within urban areas and therefore will not gain desirable IT skills.[396] This makes these individuals less likely to have access to, the ability to work with, or benefit from emerging technologies such as AI. This lack of access to, and understanding of, technology means that those within such communities are often unable to break out of this cycle, and thus the digital divide continues to deepen.

AI has the potential to add to this already existing issue; those involved with the development and deployment of the technology will likely come from areas that already

[392] Ibid n386
[393] Ibid n384
[394] K. Aruleba, N. Jere, 'Exploring digital transforming challenges in rural areas of South Africa through a systematic review of empirical studies' (2022) *Scientific African* 16 < https://doi.org/10.1016/j.sciaf.2022.e01190> accessed 30/03/2023
[395] K. Salemink, D. Strijker, G. Bosworth, 'Rural development in the digital age: A systematic literature review on unequal ICT availability, adoption, and use in rural areas' (2017) *Journal of Rural Studies* 54 360-371
[396] Ibid n384

benefit from good digital infrastructure, leaving those who are not from these areas out of the field entirely. This will also only further the issue regarding the lack of representative data in South Africa needed to accurately train AI systems. Therefore, in order for South African society to truly benefit from AI-based technologies, more must be done to minimise the existing digital divide.

There is evidence that the South African government acknowledge this issue and are intending to make some attempt to combat the divide. One example of this is the various Smart City initiatives being pursued in South Africa. At present, the focus is on the three major South African cities of Johannesburg, Cape Town and Durban.[397] Smart solutions are being explored in these cities that include methods for dealing with traffic and congestion, and more generally how technology can be used to provide better services and quality of life to those within these cities.[398] On the face of it, it seems that the development of smart cities in South Africa should not be a priority, when so many communities lack the infrastructure for basic services such as water, sanitation or electricity.[399] However, within the South African Smart Cities Framework, it is acknowledged that smart city initiatives need to respond accordingly to local challenges as opposed to following a general model that might not be suitable to South African conditions.[400]

Whilst there mightn't be specific reference within the framework to how exactly these 'local challenges' should be tackled; it is at least evidence that the South African government acknowledges that when it comes to the implementation of emerging technologies approaches taken by other nations mightn't always suit South African conditions.

There are also efforts being pursued by the South African government to train up to one million young people in AI, coding, and robotics by 2030.[401] Again, initiatives such as this might not entirely rid South Africa of the digital divide, but it will go some way in minimising it. By increasing the number of young people with technical skills and education, the future AI workforce will be diversified, meaning we are less likely to see instances of bias and discrimination from occurring.

Promoting 'local AI' is something also encouraged by Smart Africa's 2021 blueprint, Smart Africa being an AI programme lead by South Africa and the German Agency for International

---

[397] International Trade Administration, 'South Africa – Country Commercial Guide' (*trade.gov*, 2023) < https://www.trade.gov/knowledge-product/south-africa-information-technology > accessed 30/03/2023
[398] Ibid
[399] Cooperative Governance, Republic of South Africa 'A South African Smart Cities Framework' (*cogta.gov.za*, 2021) < https://www.cogta.gov.za/cgta_2016/wp-content/uploads/2023/01/Annexure-A-DCoG_Smart-Cities-Framework.pdf> accessed 30/03/2023
[400] Ibid
[401] Ibid n397

Cooperation.[402] The blueprint states that a promising way forward for South Africa and other African nations is via developing local AI, which means South Africa will be able to avoid "dependencies from international platform monopolies in the field of data provision, data processing and AI solutions".[403] This combines the efforts discussed above to develop AI-literacy in South Africa, and also potentially decreases the likelihood of biased data sets being used within AI in the nation as more diverse and representative data sets could be developed and relied upon.

3.5.3 How do the initiatives here embed the principles of accountability, transparency, and non-discrimination?

The three key ethical issues that we usually see associated with AI, accountability, transparency, and discrimination, still resonate within South Africa but there are some additional ethical issues identified here that are unique to the South African nation, these being; a shortage of data that is truly representative of society, lack of understanding regarding the technology itself at both public and government level, and that the technology itself might well deepen existing inequalities.[404]

This tells us several things, firstly that those three key ethical issues (accountability, transparency, and non-discrimination) are mostly universal, but South Africa faces additional issues that are predominantly "socio-technical" in nature.[405] This means that in comparison to other nations, any AI policy or governance measures taken in South Africa must attempt to deal with these unique factors in addition to, and most likely before, the already well-established ethical issues we see prioritised by other nations and regions.

To aid in addressing these factors, it will be valuable to consider AI from selection of critical lenses (e.g., a critical race lens, a gendered or social class lens) in order to identify the root causes of these problems, and to find potential methods for minimising them (as considered in Chapter Two of this thesis). South Africa therefore offers a unique perspective on the issues that some nations, particularly those within the global south, might face when it comes to governing AI when compared to the approaches taken by nations in the global north. It is crucial that we understand these issues to better promote international harmonisation, as far as is possible, when it comes to the regulation of artificial intelligence.

---

[402] Diplo, 'Artificial Intelligence in Africa: Continental policies and initiatives' (diplomacy.edu, 2022) < https://www.diplomacy.edu/resource/report-stronger-digital-voices-from-africa/ai-africa-continental-policies/> accessed 01/04/2023
[403] Ibid n399
[404] Ibid
[405] Ibid

3.5.4 Conclusions

Despite the lack of regulation and clear-cut AI policy at present, South Africa is still deemed as one of the top five placed African governments in terms of AI readiness as of 2019.[406] Therefore, it will be interesting to see how the nation approaches AI governance in coming years, and which specific AI-related issues they will choose to prioritise given the prominence of the socio-technological issues discussed in this section. It might very well be the case that the South African government chooses to base their AI regulation upon a framework established by another nation or region, perhaps choosing a sector-by-sector approach similar to the US and UK, or even a more blanket-style regulation as favoured by the EU. Nonetheless, South Africa highlights some of the unique issues that nations within the global south face when it comes to implementing and regulating emerging technologies such as AI.

3.6 Regulatory Strategies in Egypt

To complete this Chapter, the final analysis here will be on Egypt and its regulatory approaches to AI. Similar to South Africa, Egypt was chosen to be examined within this thesis as it offers a unique perspective of a developing country that spent a long time deciding whether or not it needed an AI strategy. Ultimately Egypt decided that they wanted to be a part of the conversation on AI development, and that they wanted AI to work for them and not them for it.[407] As a result, Egypt launched their AI strategy which will be considered in this section.

3.6.1 Egypt National Artificial Intelligence Strategy

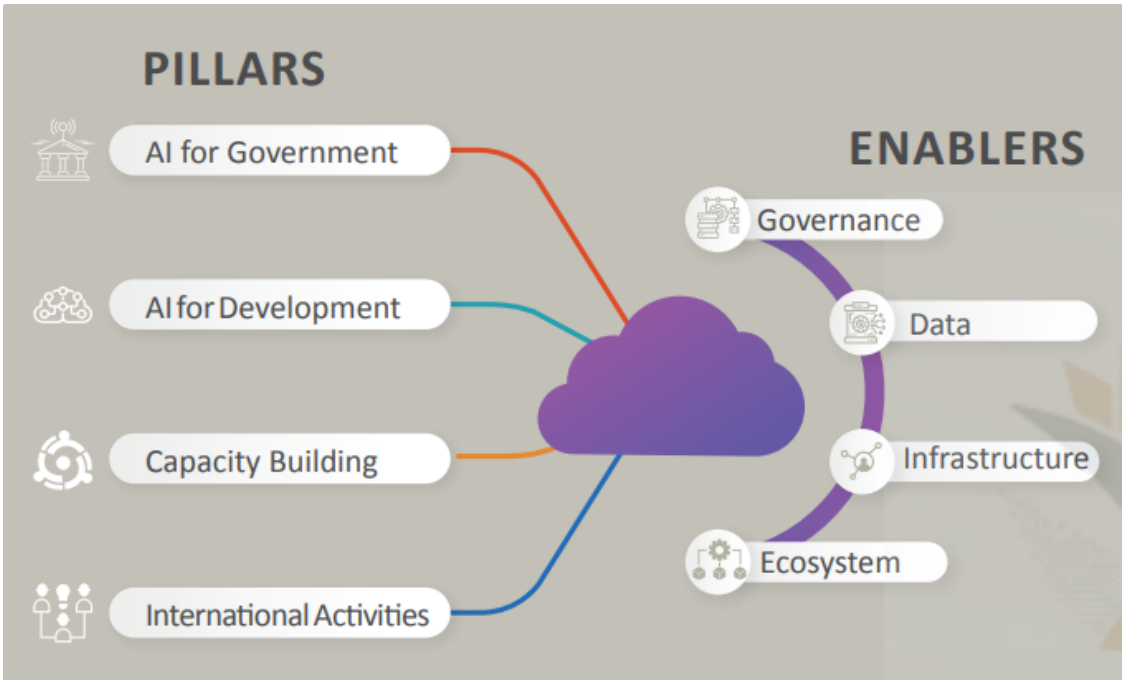
Egypt's National AI Strategy (herein referred to as the Strategy) was published in July 2021 and has two primary objectives; to achieve Egypt's development goals by exploiting AI, and to take part in regional and international conversations on AI governance.[408] These goals therefore set out to establish an AI industry within Egypt, meaning that in a similar vein to most other regional AI strategies examined so far in this thesis, the Egyptian Strategy focuses significantly on using AI as a tool for development. Therefore, it is questionable as to how far the Strategy tackles some of the commonly acknowledged legal and ethical issues

[406] Oxford Insights, 'Government Artificial Intelligence Readiness Index' (2019) < https://africa.ai4d.ai/wp-content/uploads/2019/05/ai-gov-readiness-report_v08.pdf> accessed 30/03/2023
[407] S. Radwan, 'Egypt's Ai strategy is more about development than AI' (oecd.ai, 2021) < https://oecd.ai/fr/wonk/egypt-ai-strategy> accessed 06/04/2023
[408] Ibid n249

posed by AI, due to the document prioritising the use of AI as a means for national economic development instead.

The strategy can therefore be split into four distinct pillars, which detail areas in which AI can be used to encourage development across the nation, and four enablers which enable the Strategy to succeed.

Figure 5: Pillars and enablers of the Egyptian AI Strategy[409]



Source: The National Council of Artificial Intelligence, 'Egypt National Artificial Intelligence Strategy' (2022) < https://mcit.gov.eg/Upcont/Documents/Publications_672021000_Egypt-National-AI-Strategy-English.pdf> accessed 26/03/2023

The pillars each cover a wide variety of AI-related goals. AI for Government places specific focus on increasing the effectiveness of public services for citizens.[410] AI for Development focuses on using AI within agriculture, healthcare, economic planning and within finance and banking.[411] Capacity Building focuses on raising public awareness of AI and upskilling the population so Egyptian citizens can develop and work alongside AI.[412] And International Activities focuses on developing the role that Egypt plays in AI development and governance at an international level.[413] Therefore, it is clear that Egypt intends to leverage AI as a tool for its own domestic development, which could prove to be very fruitful.

[409] Ibid n249
[410] Ibid n249 p. 26
[411] Ibid n249 p. 29-35
[412] Ibid n249 p. 36-43
[413] Ibid n249 p. 44-45

According to Radwan, the most important pillar within the Strategy is the capacity building pillar as it essential for achieving all other goals set out within the Strategy.[414] This is quite similar to the goals that we have seen prioritised by the South African government, in that it is necessary to ensure that all generations of citizens are equipped with the skills and knowledge to both develop and work with AI-based technologies as a foundation for any AI governance measures. This is also akin to the UK approach in which we see education featuring as part of the National AI Strategy.

As a result, Egypt is currently pursuing a number of pilot schemes in order to determine the best routes for improving general AI awareness and AI-literacy, one of which is a scheme targeted at high school students and will see them get hands-on experience in implementing AI via small projects.[415] Similarly, since the inception of the Strategy Innovation Hubs have been established across six universities in Egypt, with the aim being to establish connections between universities and local communities, enhancing tech-related education and encouraging entrepreneurship.[416]

Despite the potential benefits of the approach to using AI to help drive development within Egypt, there are some potential concerns that should be addressed. For example, the AI for Government pillar predominantly references using AI to improve government services and therefore increase efficiency. Many of the key areas that have been identified for how AI can support government operations involve using AI to automate manual tasks and decision-making.[417] Whilst AI does have the capacity to increase efficiency in the delivery of government services, using AI in this way also has the potential to cause significant negative impacts. There must be adequate safeguards in place to prevent instances such as that which occurred in South Africa, regarding the automation of distributing government grants, from occurring in Egypt also.

Perhaps enhancing existing data protection laws within Egypt such as the recent Law on the Protection of Personal Data 2020,[418] to better protect the data rights of Egyptian citizens when it comes to automated decision-making might go some way in ensuring that instances such as the above are avoided. Nonetheless, there is limited information provided in the

[414] Ibid n407
[415] Ibid n407
[416] Ministry of Communications and Information Technology, 'Boost your Business - Creativa Innovation Hubs' (mcit.gov.eg, 2023) < https://mcit.gov.eg/en/Innovation/Boost_your_Business/Creativa_Innovation_Hubs> accessed 20/05/2023
[417] Ibid n249 p. 28
[418] The Law on the Protection of Personal Data ('the Data Protection Law') issued under Resolution No.151 of 2020

Strategy as to how the Egyptian government intends to avoid the negative impacts that often result from automated decision-making at government level.

Further to this it is vital to acknowledge that with wide-spread AI use, as Egypt intends via the Strategy, the issue of a lack of representative data to train AI as discussed in the previous section on South Africa applies to Egypt also. In order for Egypt to truly reap the benefits of AI, and to 'make AI work for them' as the government describes, there needs to be meaningful work done to make AI systems deployed in Egypt representative of the communities they serve. This is the case for any type of AI, whether we are referring to automated decision-making algorithms, or facial recognition tech that is typical developed in the global north.

The efforts by the Egyptian governments to enhance AI-literacy via the third pillar of the Strategy may go some way to achieving this representation, or at least improving it, however this will not happen overnight. It is therefore something the Egyptian government need to be aware of as they pursue their pro-AI approach.

3.6.2 How do these initiatives embed the principles of accountability, transparency, and non-discrimination?

As above, there are a number of capacities in which the Strategy incentivises the use of AI which give rise to ethical considerations, primarily the likelihood of these AI applications causing discrimination. Whilst the Strategy itself gives little indication as to how key ethical principles will be incorporated into Egyptian AI governance, in April 2023 the National Council for Artificial Intelligence announced the new Egyptian Charter for Responsible AI.[419]

The Charter prioritises five key ethical AI principles including human-centred design, transparency, justice, accountability, and security. Therefore, the Charter directly establishes both transparency and accountability as key components to the Egyptian regime. Whilst non-discrimination isn't specifically mentioned as a key principle within the Charter, the principle of fairness here can be read as achieving the same goals; to ensure that no one is harmed by the implementation of AI, and that individuals have a means of redress in the instance that they are harmed by AI.

Therefore, whilst the initial Strategy gives little attention to the ethical dimensions of AI, the supplementary Charter for Responsible AI that has followed almost two years after the Strategy provides this ethical context. This is method that nations around the world could

[419] National Council for Artificial Intelligence, 'Egyptian Charter for Responsible AI' (mcit.gov.eg, 2023) <https://mcit.gov.eg/en/Media_Center/Press_Room/Press_Releases/66939#:~:text=Egypt%20was%20also%20the%20first,Accountability%2C%20and%20Security%20and%20Safety.> accessed 20/05/2023

adopt if for example their existing AI initiatives, strategies and frameworks do not give enough attention to ethical AI principles such as accountability, transparency, and non-discrimination. Using a supplementary document like the Egyptian Charter for Responsible AI means that ethical considerations can be given deeper analysis and attention than they might be afforded in a typical National Strategy.

Once again, it is however worth noting that the Charter is not legally binding which limits its enforceability. Although, the Egyptian government do acknowledge this by stating that the Charter is meant to act as a 'soft launch' which will empower citizens to expect ethical AI as and when it is deployed in Egypt.[420] Therefore, the document acts as a further signal that represents Egypt's preparedness for AI investment and adoption.[421]

3.6.3 Conclusions

The Egyptian AI Strategy and Charter for Responsible AI are significant milestones that signify Egypt's desire to become an integral part of the international conversation on AI development and governance. As pointed out in this section, Egypt is predominantly using these two documents as signals to the rest of the African and Arab regions that they are ready to play an active role in the global AI discussion, but these documents also signify that Egypt's primary intention is to use AI as a tool for development. This is an interesting and somewhat unique approach taken by this nation, and it will be interesting to see how these efforts come to fruition going forward.

3.7 Conclusions

This chapter highlights the approaches taken by a variety of nations including the UK, the US, China, South Africa, and Egypt. As demonstrated, the regulatory initiatives, strategies and frameworks promoted by each of these states are relatively varied, with some similarities flagged throughout. The benefit of conducting such an analysis is to establish what works well within the AI regulatory landscape, and what does not, and helps to identify common themes that exist between nations despite there being a mixture of legally binding and non-binding measures considered here.

Chapter Four builds on this analysis by considering regional and international regulatory frameworks and strategies on AI, including an examination of the EU, Africa, and the United Nations. This chapter therefore tackles research question two, specifically how well equipped are current legal instruments, proposed legal instruments, strategies, and frameworks in dealing with the issues posed by AI?

[420] Ibid
[421] Ibid n419

Chapters Five and Six of this thesis will take these analyses into account and proceed to consider what the ideal regulatory response might look like. Taking on board the issues that come along with regulating technology, general issues in creating legislation, and how to improve upon the proposals presented within this Chapter and the next.

3.7.1 Recommendations to nations or states on the regulation of AI

For clarity, it is worth presenting a few key recommendations to those wishing to regulate AI domestically, as is the case for the jurisdictions examined in this Chapter. Firstly, there are a number of key points to take away from this analysis of jurisdictions so far:

- Most nations appear to favour sector-by-sector regulation as opposed to imposing a blanket-style regulation.
- Trustworthiness is a key aim, if not the central aim, of most regulatory strategies. This means to some extent most nations give attention to ethical AI principles such as accountability, transparency, and non-discrimination.
- Balancing technological innovation with the need to regulate is still proving to be a significant task.
- It appears to be universally agreed that AI is in need of regulation, however, most proposals and legislative attempts lack a sufficient amount of detail to make meaningful impact (with the possible exception of China's recent efforts).

The following recommendations are made:

Nations should carefully consider what their interests in AI are prior to pursuing regulation. As demonstrated within the analysis of both South Africa and Egypt, these nations have identified some unique barriers to AI development and deployment in their states, and therefore are considering ways in which they can leverage AI to benefit them specifically. This is rather different from the approach being taken by countries such as the UK and US in the global north, in which nations are typically focused on positioning themselves as AI 'superpowers'. AI regulation is going to be necessary across the globe, but each nation should first carefully consider what exactly they want from AI, what the potential barriers might be in their 112jurisdiction, and how they can leverage AI to better suit their needs.

Building on this recommendation, it is suggested that within policy documents, strategies and frameworks published by governments to signal their intentions regarding AI governance, there should be more explicit incorporation and acknowledgement of key ethical AI principles. Incorporating key principles such as transparency, accountability, and non-discrimination within AI at the development stage is one of the easiest ways to produce reliable AI. These three principles in particular go hand in hand, and with the implementation

of one of these principles (typically transparency) the other two can be achieved. This thesis will explore precisely how transparency might be embedded within regulation within Chapter Five of this thesis.

There are a plethora of ethical guidelines that can be referenced and used to this end, e.g. the HLEG Ethics Guidelines and SHERPA principles,[422] the Future of Life Institute Asilomar AI Principles,[423] and the UNI Global Union Top 10 Principles for Ethical AI.[424] This list is by no means exhaustive, but indicative of the number of ethical guidelines available to nations for use in AI governance strategies and frameworks. An approach that achieves the same means is the one taken by Egypt who published their own ethics charter to supplement their AI strategy.

Nations should invest in education and upskilling their populations in order to promote AI-literacy. This is a common theme amongst most of the nations considered here and will be necessary regardless of where the nation is situated, e.g., global north or south. Populations need to be equipped with the right skillset to thrive in an increasingly AI-dominated world. In some nations, e.g., within South Africa and Egypt, pursuing this particular goal is going to be the cornerstone of any future AI governance measures.

As identified earlier in this section, it appears that there is a preference amongst nations to pursue a sector-by-sector regulatory approach to AI. It is therefore recommended that if nations do wish to pursue this model that they consider how exactly this approach will be achieved, e.g., providing detail such as which regulators will be relied upon, if new agencies need to be created when will this happen, what about industries that overlap, who will be responsible for their regulation etc. Clarity in these documents is beneficial for a number of stakeholders including those within government, industry, and academia.

[422] See page 29 for further detail
[423] Future of Life Institute, 'The Asilomar AI Principles' (futureoflife.org, 2017) < https://futureoflife.org/open-letter/ai-principles/> accessed 20/05/2023
[424] UNI Global Union, 'Top 10 Principles for Ethical Artificial Intelligence' (thefutureworldofwork.org, 2017) < http://www.thefutureworldofwork.org/opinions/10-principles-for-ethical-ai/> accessed 20/05/2023

Chapter Four

Comparing Regional and International Frameworks and Strategies on AI: The EU, Africa, and the United Nations

4.1 Regulatory Strategies in the European Union (EU)

The regulation of AI has been firmly placed within the EU legislative agenda for quite some time, with early proposals from the European Parliament in 2017 proposing to create a form of electronic personhood for artificially intelligent robots, and motions to install a more robust EU-wide AI legislative agenda since then.[425] The overall ambition of the EU in this sense is to create trustworthy AI and to promote excellence and best practice within this area.[426] Certainly, creating and installing a system that ensures that all AI-systems that are developed and deployed within the EU have the principles of trustworthiness embedded within would be incredibly beneficial, and would likely have positive global impact beyond EU member state countries.[427]

However, the question remains as to how and if this high-level goal will be achieved. Therefore, this section will investigate the EU AI regulatory timeline in order to fully consider the various proposals, recommendations and schemes adopted within the EU, and the effectiveness of these approaches. In assessing this legislative timeline, the proposed EU Artificial Intelligence Act will be considered in some detail alongside its accompanying measures such as the digital sector package of provisions and amendments.[428]

4.1.1 The Proposed Artificial Intelligence (AI) Act

The formal proposal for the EU AI Act was published by the European Commission in April 2021. Initial analysis of the draft legislation brings to light a number of noteworthy points and questions that may prove to have significant impact upon the global AI regulatory landscape. For instance, there are some significant implications that may result from the introduction of this legislation as it stands, such as its need to be read alongside other legislative components in order to be fully functional, or its lack of definitions for crucial terms, whilst in some provisions the proposal appears to be wholly ineffective.

---

[425] European Parliament Committee on Legal Affairs, 'Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) 31.5.2016

[426] European Commission, 'A European approach to artificial intelligence' (digital-strategy.ec.europa.eu 2022)< https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> accessed 01/02/2022

[427] ibid

[428] European Commission, 'The Digital Services Act package' (Digital-Strategy.ac.europa.eu, 2022) < <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> accessed 03/02/2022

Therefore, an in-depth analysis of this proposal is warranted in order to identify provisions and approaches that might potentially work well to regulate AI, to identify those that could be improved upon, and to ultimately get a better idea as to what the proposed regulatory landscape may look like in this space in the very near future. This type of analysis will no doubt prove useful to countries such as the UK in forming their future regulatory framework which at present is quite minimal, and more importantly to help paint a better picture as to what the 'ideal' regulatory approach to AI regulation may look like; this being the central aim of this thesis.

As discussed briefly above, the AI Act is intended to operate alongside other similar EU provisions created to govern the continuously emerging digital tech space including the draft Digital Services Act[429], the draft Digital Markets Act[430], draft Machinery Regulation[431] and the draft Data Governance Act[432], and also potentially updated and more tech friendly product liability legislation (these measures will be referred to in this thesis as the digital sector package).

In addition to these measures, and in order for the Act to function in its capacity to protect individual human rights and fundamental freedoms from AI-related harms, the Act must also be read and implemented alongside key human rights instruments. The European Convention for Protection of Human Rights and Fundamental Freedoms for example provides protection for various rights such as fair trial, respect for private and family life, freedom of expression and assembly as well prohibition from discrimination.[433] Furthermore, the Charter of Fundamental Rights of the European Union may be particularly useful in this space with regards to Article 8, and the protection of personal data.[434]

Considering GDPR is also necessary in this space with specific regard to transparency, proportionality, and legality as regards data use.[435] In addition, the Law Enforcement

[429] European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM (2020) 825 final)
[430] European Commission, Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) (COM (2020) 842 final)
[431] European Commission, Proposal for a Regulation of the European Parliament and of the Council on machinery products (Machinery Regulation) (COM (2021) 202 final)
[432] European Commission, Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) (COM (2020) 767 final)
[433] Council of Europe, European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950, ETS 5
[434] European Union, Charter of Fundamental Rights of the European Union, 26 October 2012, 2012/C 326/02
[435] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1

Directive of 2016 is also applicable in this space in certain circumstances.[436] Therefore, the Act alone will not have all the relevant provisions to address the human rights issues AI presents us with yet reading it alongside these existing provisions will help.

With regards to the updating and amendment of product safety and liability regulations, the wording of the proposed AI Act in itself is largely lifted from a decision of the European Parliament and Council regarding product safety.[437] Upon closer analysis, the proposed Act hopes to achieve a number of the same aims as the newly proposed product safety regulations as published by the European Commission in the summer of 2021.[438] Therefore, an initial observation might be that whilst the proposed AI Act is a relatively contemporary device it cannot be read or interpreted as a standalone regulatory measure, it must be read in conjunction with the various legislative components of the digital sector package and various human rights provisions.

The proposed Act also builds upon policies and working documents published by the European Commission and its High Level Expert Group on AI, including the 2019 Ethics Guidelines for Trustworthy AI[439] and its 2020 White Paper on excellence and trust in AI.[440] As is relatively clear from the names of these documents, they primarily consider what steps we can take to promote trustworthy AI, and so therefore they focus considerably on the ethical concerns regarding AI as opposed to the concept of 'lawful AI' that is promoted within the proposed AI Act.[441] As a result, at present we do not have a sufficiently detailed enough understanding as to what the central component of the legislation, 'lawful AI', actually is which is leaves us with a degree of legal uncertainty. This is something considered in detail by the LEADS Lab at the University of Birmingham, who develop the

[436] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA

[437] M. Veale, F. Z. Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) *Computer Law Review International,* 4 97-112

[438] Think Tank, European Parliament, 'General product safety regulation' (europarl.europa.eu, 2021) < https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698028> accessed 10/02/2022

[439] European Commission High Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019)

[440] European Commission, *Artificial Intelligence – A European approach to excellence and trust* (White Paper (COM 2020) 65 final)

[441] N. Smuha, E. Ahmed-Rengers, A. Harkens, W. Li, J. MacLaren, R. Piselli, K. Yeung, 'How the EU can achieve Legally Trustworthy AI: A response to the European Commission's proposal for an Artificial Intelligence Act' (2021) *LEADS Lab @ University of Birmingham*

concept of 'lawful AI' in their paper on the topic and what this might actually mean within the context of the proposed legislation.[442]

The proposed AI Act is on the whole an interesting piece with great promise. It is therefore worthwhile considering what works well within the proposal, and which areas need to be reconsidered or considered further for the Act to be truly functional and effective.

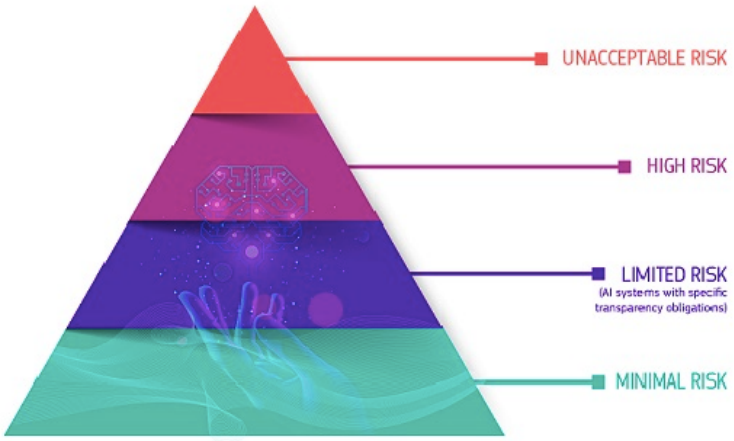
The proposal itself is a powerful expression made by the European Commission that supports the notion that in order to safely govern the development and use of AI, we must have legally enforceable measures; guidelines, frameworks and voluntary standards are not enough alone. This perspective is welcomed, and it is also encouraging to see that alongside this robust statement the Commission are keen to ensure that we not only regulate AI but that we continue to promote its use.[443] This is one of the key issues flagged regarding the regulation of modern technology within this thesis, creating legal measures that are future-proofed and flexible is no easy task, but finding the balance here is going to be vital to the success of any modern governance measure.

The proposed legislation applies to public and private developers where an AI system is placed on the market in the EU or where it affects people in the EU (similar to GDPR). The Act therefore creates a number of new obligations on providers and developers of high-risk AI systems, with many activities being identified as high-risk, for example using AI within justice, immigration, and legal settings. These risk categories and new obligations will be explored in more detail shortly,

The legislation also proclaims to be based upon and to uphold basic fundamental rights and EU values, therefore the rules contained within the proposed Act will be human-centric.[444] By creating rules that are first and foremost focused upon the protection of basic human and fundamental rights, this will likely provoke faith in the fairness and lawfulness of future AI applications. Again, finding the correct balance between the protection of fundamental rights, public security, and promoting the development and use of AI is going to be a challenge throughout the legislative process.[445]

[442] Ibid
[443] M. Anzini, 'EIPA Briefing 2021/5, The Artificial Intelligence Act Proposal and its implications for Member States' (eipa.eu, 2021) < https://www.eipa.eu/publications/briefing/the-artificial-intelligence-act-proposal-and-its-implications-for-member-states/> accessed 10/02/2022
[444] Ibid n8
[445] Ibid n443

Figure 6: The risk-based approach[446]



Source: European Commission, 'Regulatory framework proposal on artificial intelligence' (2022) https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai, accessed 11/02/2022

Additionally, the proposed Act presents an organised and clear method of regulation; in the opinion of the author, the structure of the risk-based approach of the Act is the strongest aspect of the overall proposal, however, it is not without shortcomings. The proposed Act defines four categories of risk types associated with AI use, the lowest being minimal or no risk, the next being limited risk, then high risk and finally unacceptable risk. As demonstrated in the above figure promoted by the European Commission in various press releases and publications relating to the Act, the general structure is easy to follow for those perhaps unfamiliar with AI and its associated risk factors.

Some of these risk categories are better defined than others though. The high-risk category contains a good amount of explanation regarding the type of AI application that may fall within the category. For example, this category may include any of the following:[447]

- Critical infrastructures (e.g., transport), that could put the life and health of citizens at risk
- Educational or vocational training, that may determine the access to education and professional course of someone's life (e.g., scoring of exams)
- Safety components of products (e.g., AI application in robot-assisted surgery)
- Employment, management of workers and access to self-employment (e.g., CV-sorting software for recruitment procedures)

[446] European Commission, 'Regulatory framework proposal on artificial intelligence' (digital-strategy.ec.europa.eu, 2022) < https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 11/02/2022
[447] Ibid

- Essential private and public services (e.g., credit scoring denying citizens opportunity to obtain a loan)
- Law enforcement that may interfere with people's fundamental rights (e.g., evaluation of the reliability of evidence)
- Migration, asylum, and border control management (e.g., verification of authenticity of travel documents)
- Administration of justice and democratic processes (e.g., applying the law to a concrete set of facts)

As demonstrated in Chapter Two of this thesis which considered various uses of AI and the arising impacts of such, most of these applications classified as high risk by the European Commission have been at the centre of various scandals relating to algorithmic bias and discrimination in the past couple of years. Therefore, addressing these applications as high risk and assigning them their own restrictive measures definitely seems like a step towards solving this ever-increasing issue.

In addition to this, the proposed Act lists various requirements and obligations that these high-risk applications will be subject to in order to be placed upon the market in the EU, these include:

- An adequate risk assessment carried out and mitigation systems in place
- The datasets feeding the system must be high quality in order to minimise risks and discriminatory outcomes
- Logging of activity to ensure traceability of results
- Detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance
- Clear and adequate information provided to the user
- Appropriate human oversight measures to minimise risk
- High level of robustness, security, and accuracy in the system

Therefore, applications that are somehow categorised as lower risk will not be subject to the same obligations and restrictions as the previously listed applications, demonstrating the real mix of risk-based thinking and rulemaking being undertaken by the Commission in the preparation of this Act.[448] Yet, the question remains as to how we fairly decide which

[448] T. Mahler, 'Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal' (2022) *Nordic Yearbook of Law and Informatics* < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4001444> accessed 10/02/2022

applications are to fit within each category. For example, the unacceptable risk category is rather vague and we are told that it will include "all AI systems considered a clear threat to the safety, livelihoods and rights of people."[449] These applications will be banned, and might include anything from applications used for "social scoring by governments to toys using voice assistance that encourages dangerous behaviour".[450] However, the two applications used here as examples by the Commission could quite easily fall within the high-risk category, so it becomes unclear as to what truly sets these two categories apart.

On the other hand, the differences between the applications that may fall within both the limited risk and minimal to no-risk categories are also difficult to differentiate. For the limited risk category, we are told that this includes "AI systems with specific transparency obligations."[451] Again, this is rather vague as arguably all AI systems should be subject to transparency obligations no matter the scale of their intended application and potential resulting impacts. In addition, we are given the example of "using AI systems such as chatbots" and that in this instance "users should be aware that they are interacting with a machine so they can take an informed decision to continue or step back".[452] Again, this should be the case for all AI applications; users should be clearly informed that they are interacting with an AI-based system in whatever function that might be (whether it is directly or via automated decision-making for example) and therefore have the choice to at least receive information about the system and its functionality. Therefore, this category does not seem all that functional, or at least our present understanding of this category and its requirements and obligations is most definitely limited.

Regarding the most basic and risk-averse category 'minimal or no-risk', we are told that "This includes applications such as AI-enabled video games or spam filters."[453] Once again, it is challenging to see how this category is all that different from the limited risk category, it would therefore be useful to know what really sets these two types of AI application apart as at present it is quite unclear. The Commission then go on to assert that "The vast majority of AI systems currently used in the EU fall into this category."[454] However, the applications listed as potentially falling within the proposed high-risk category are so far-reaching and broad in scope that is questionable as to whether the statement that the vast majority of AI systems within the EU will fall within the lowest risk category is all that accurate.

[449] Ibid n446
[450] Ibid n446
[451] Ibid n446
[452] Ibid n446
[453] Ibid n446
[454] Ibid n446

Therefore, this leads us to question the actual purpose and logic behind the risk-based approach adopted within the proposed Act. Upon first reading, the chosen risk-based approach could be seen as an appropriate response to the identification by the Commission that different AI applications have different levels of risk associated with them; therefore, a risk-management methodology would be more suitable as opposed to a blanket-style legislation which might be seen as counterintuitive and futile in this space.

Yet, when considering the risk-based approach presented within the Act in further detail, as put by Mahler, it appears as though the intention here mightn't actually be to manage risk but to 'ensure legislative proportionality'.[455] This is perhaps reflected in the rather vague descriptions given of the AI applications that might fit within the four categories given (namely limited risk and unacceptable risk), meaning that practically the obligations and restrictions that may eventually be attached to these categories will likely not be robust enough to have meaningful impact and may function ineffectively.

Earlier it was stated that in the opinion of the author, the risk-based approach adopted within this proposal was possibly the strongest aspect of the proposed Act, but whilst this approach has merit, it is not without room for improvement. A risk-based approach is conceivably the most logical regulatory approach as it has the most potential in achieving the balance between ensuring safe and ethical AI whilst also protecting and promoting innovation. However, it is suggested that the four risk categories proposed within the AI Act are not developed to a point in which they would actually be functional and effective, and so it therefore falls short.

Whilst the proposed Act appears to be the most logical plan currently in development for the governance of AI, there are still several areas for improvement within the proposal. By addressing these shortcomings, the proposed Act has the potential to be a ground-breaking step forward in the regulation of AI. At present however, the potential effectiveness of the Act is diminished.

There are three central areas of improvement that will be considered in this thesis. Firstly, the proposal does not fully recognise or acknowledge the actual harm caused by AI (e.g., threats to fundamental rights etc.). This therefore calls into question further the functionality of the risk-based approach employed by the Act, similarly the criteria for including an application within a given risk category should be more clearly defined within the Act. Secondly, it is not clear how future proof the Act itself will be, for example, we are presented with a list of 'high-risk' AI types by the Commission but to be truly functional, this list should

[455] Ibid n446

and will have to be amendable, yet this isn't addressed within the Act. Thirdly, the choice to approach AI in a similar vein to that of product safety regulations should be reconsidered; AI can be seen to have its own lifecycle and is therefore distinctly different from singular products and services and should be treated as such.[456] Resultingly, it is questionable as to whether the chosen approach is all that suitable.

The first point raised here is arguably the most notable shortcoming of the Act; namely, that the Act should do more to recognise the actual harms caused by AI. At present, the regulatory measures introduced by the Act claim to uphold basic fundamental rights, making the rules contained within the proposed Act human centric. The Charter of Fundamental Rights of the European Union (the Charter) enshrines within law the rights of European citizens with respect to freedoms and values upheld within the EU.[457] The Charter protects a diverse selection of rights and freedoms including the right to human dignity, protection of personal data, and the right to a fair trial; the rights listed here are just some of those that that are potentially at risk of breach by the use of AI.

Yet the Act does very little to recognise the actual harm caused by AI and as such, the threats that this poses to our fundamental rights (for example the right to private family life, or the right to freedom of association and assembly).[458] For instance, there are frequent references made throughout the proposal to fundamental rights, however these references are relatively vague. There is also some reference to the proposed provisions addressing opacity, bias, and unpredictability in order to ensure compatibility within fundamental rights, yet there is little detail available regarding how this will be achieved (aside from assigning AI likely to cause this type of damage to the high-risk or unacceptable risk categories).[459]

To resolve this shortcoming and to strengthen the proposed Act, it is suggested that amending the risk-based approach to more specifically address the actual and potential harms likely to result from the applications contained in each category, would more effectively protect our fundamental rights. For example, within the high-risk category, as opposed to solely listing the types of applications that usually conflict with our human rights, it is suggested that listing the rights with the highest risk of violation from AI-based systems would allow for an Act that is wider in scope, more flexible, whilst still workable. Therefore, a

[456] A. Circiumaru, 'Three proposals to strengthen the EU Artificial Intelligence Act' (adalovelaceinstitute.org, 2021) < https://www.adalovelaceinstitute.org/blog/three-proposals-strengthen-eu-artificial-intelligence-act/#:~:text=The%20European%20Commission's%20Artificial%20Intelligence,in%20law%20enforcement%2C%20education%20and> accessed 15/02/2022
[457] Ibid n71
[458] Ibid n71
[459] Ibid n8 1.1

risk-based approach can still be utilised within the Act, whilst ensuring that the overall aim of the Act taking a human-centric approach to regulation can safely be achieved.

In addition to the flaws within the current risk-based approach, the Act could make further reference to those ultimately affected by adverse AI, namely the public. With the Act aiming to improve the safety of society at large from AI-based threats, it could do more to empower the public in ensuring that their AI-based experiences are fair and in line with those exact legal measures. In order to encourage developments in the Act to this tune, EDRi (European Digital Rights), a civil society group working to improve digital rights across Europe, are urging the commission to amend the Act to enable those affected by harmful AI to seek remedies via granting individuals specific rights regarding the uses of AI.[460] Motions for amendments such as this one are widely supported amongst similar groups, and again, strive to help the proposed Act to achieve its human-centric aims.

The Proposal places obligations upon those involved in the AI supply chain, these actors are labelled by the Act as 'providers', which includes any person that develops or has an AI developed with the view to putting it on the market in the EU or putting into service.[461] The Act also places obligations on 'users', which are defined within Article 3(4) of the Act to mean any natural or legal person "using an AI system under its authority", this might include a bank putting in a automated loan approval scheme, or a business using a hiring algorithm'.[462] There is potential for confusion here regarding the choice of the term 'user' to define this person. As explained by Lilian Edwards, Professor of Law, Innovation and Society at Newcastle University, the 'user' referred to by the Act is not the ultimate 'end-user' of the system, or the data subject in terms of GDPR.[463] The Act in fact does not contain a word for the ultimate end user of the system, despite the fact that the Act exists to benefit this person.

This is a rather confusing choice of terminology which could potentially provide for legal uncertainty, especially amongst the public who may possibly look to the Act to try to better understand their legal rights regarding the use of AI systems. It also strengthens further the point made earlier; the 'end-user' could be better represented within the proposed regulations. It is therefore suggested that the Commission revisits the proposal to more closely consider how the end-user can be included within the Act and empowered by the

[460] EDRi, 'The EU AI Act and fundamental rights: Updates on the political process' (edri.org, 2022) < https://edri.org/our-work/the-eu-ai-act-and-fundamental-rights-updates-on-the-political-process/> 10/03/2022
[461] Ibid n8, Article 3(2)
[462] Ibid n8, Article 3(4)
[463] L. Edwards, 'The EU AI Act: a summary of its significance and scope' (adalovelaceinstitute.org, 2022) <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf > accessed 20/05/2023

regulations, as opposed to being treated as merely bystanders. One way in which this could be achieved is by including provisions within the Act that grant end-users with the ability to report AI harms. This reporting mechanism should be centralised and uniform, avoiding the difficulty that would otherwise inevitably arise whereby end-users have to tolerate the differing reporting mechanisms made available by various AI manufacturers.

In a similar vein, the proposed Act seems to be rather static in its composition meaning it may be difficult to amend in the near future as and when necessary. Chapter Five of this thesis considers in some detail the need for tech-based regulations to be dynamic and future-proof due to the everchanging nature of modern technology. Yet, the provisions within the AI Act seem rather stagnant. Therefore, the Commission need to ensure that adequate measures are put in place to make the various lists within each of the risk categories amendable. Indeed, heeding the suggestion made earlier within this thesis, to explicitly address the most 'at-risk' fundamental rights within each of the risk categories, this may in fact go some way in providing the flexibility and adaptability needed to create a truly future-proofed legislation. By detailing the risks based upon the potential rights that may be violated, this leaves room for newly developed technology with sophisticated abilities to fall within scope of the legislation without having to be explicitly listed within the Act.

Finally, the last shortcoming identified within the proposal is a salient point raised by Circiumaru of the Ada Lovelace Institute; the proposal seems to use a product safety regulation approach to tackle issues related to AI harms, yet AI is distinctly different from most other types of technologically sophisticated product.[464] The very nature of AI, the fact that most sophisticated AI is developed to 'think for its self' in such a way that it may learn from its environment, learn to identify patterns etc, means that the AI-system at the point of deployment may be quite different to the one a couple of weeks or months into its deployment. Therefore, applying the same product safety-style regulations that we usually apply to a product that stays relatively unchanged following deployment, to a type of technology that by nature is continually evolving may not be the most effective way to govern AI.

It is therefore proposed that the Commission considers the ever-changing quality of AI and reflect this within the proposed Act. Perhaps a solution suggested earlier in this section may go some way to address this issue; by incorporating measures within the Act that allow end-users to report harms or other undesirable AI activity, we are empowering those most likely to suffer from the harms. In this sense, we have a reporting mechanism that can help to notify manufacturers and other actors within the AI supply chain to the unintended

[464] Ibid n456

consequences resulting from their systems. Although, it is worthwhile flagging that this is a rather reactive solution as opposed to a preventative one, and preventative provisions would be much more desirable.

In a similar vein to this suggestion, the EU Commission have attempted to address the harm caused by AI to an extent via the recently proposed AI Liability Directive.[465] However, this proposal, again, does not go far enough. This directive would attempt to address harm caused by AI in a few ways; firstly, by lowering evidentiary hurdles for victims of AI harm allowing them to bring civil liability claims more easily. Secondly, the directive would make it easier for claimants to mandate disclosure of evidence concerning 'high-risk' systems, yet little information is offered as to how this would work in practice; there are clear problems with the vagueness of the risk categories proposed within the original AI Act to begin with which would likely be exacerbated via the directive. Furthermore, the proposed directive does not fully consider the issues pertaining to mandating transparency within 'high-risk' AI systems, e.g., how would the directive approach an algorithm that is protected as a trade secret similar to that within the COMPAS tool?[466]

Most importantly, this proposed directive does not impose any new legal obligations[467], and is reactive by definition as opposed to being preventative. Even if we can somehow overcome the confusion regarding the risk-categories within the proposed AI Act that are now intrinsically tied to this newly proposed directive, the provisions within the directive are only concerned with the aftermath of harm caused by an AI, as opposed to addressing the root cause and attempting to minimise risk of harm occurring entirely. Therefore, whilst this proposed directive might go some way in improving the opportunities for citizens to seek recourse when harm has occurred, its impact will be limited. This is in addition to the fact that the directive is tied to the proposed AI Act (e.g., terminology and concepts have been translated into the directive), and so any of the Act's limitations as discussed within this section, will also apply to the directive.

[465] Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) COM(2022) 496 final
[466] Ibid n59
[467] A. Gesser, R. Maddox, A. Gressel, F. Colleluori, T. Lockwood, M. Pizzi, 'The EU AI Liability Directive Will Change Artificial Intelligence Legal Risks' (debevoisedatablog.com, 2022) < https://www.debevoisedatablog.com/2022/10/24/eu-ai-liability-directive/> accessed 01/12/22

4.1.2 How does the AI Act embed the principles of accountability, transparency, and non-discrimination?

As discussed already in this section, the EU AI Act builds upon several existing AI initiatives, for example the 2019 Ethics Guidelines for Trustworthy AI published by the European Commission and its High-Level Expert Group on AI. [468] Therefore, the EU AI Act grew from a consideration of key ethical AI principles. However, after looking at the Act in close detail via this analysis there is room for principles such as accountability, transparency, and non-discrimination to be better included within the Act. Primarily, this could be achieved via the risk-based approach adopted within the Act; whilst this approach has its merits, it could be better executed by clearly identifying the potential harms that arise from the different categories of AI application, and the resulting at-risk rights (as proposed earlier in this section). Further clarity could also be provided in the likes of the recent AI Liability Directive, in which transparency is mandated for 'high-risk' systems, yet the proposed directive does not fully consider or provide for how this could be achieved e.g., how would the directive approach an algorithm that is protected as a trade secret similar to that within the COMPAS tool?[469]

Therefore, the EU AI Act does refer to these key principles to an extent, the whole point of the risk-based approach to regulation being to minimise occurrence of bias and discrimination and wider harms caused by 'riskier' AI systems. However, as suggestion in this section, these principles could be more firmly embedded within the EU AI Act, and there could be further clarity provided with regards to how principles such as transparency can be achieved, especially when they are mandated.

4.1.3 Conclusions

The proposed EU AI Act is unlike anything else currently in development; it contains some well-thought-out provisions and has established the importance of creating a meaningful system for AI regulation. The proposal's strongest element is arguably its risk-based approach, which is central to its structure, and is an innovative and seemingly practical approach to AI governance.

Despite this though, there are some shortcomings within this proposal as identified in this section. These flaws, if addressed, could potentially make the proposed Act world-leading in its approach and position the EU as the central authority on AI governance on the global

[468] European Commission High Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (digital-strategy.ec.europa, 2019) < https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 04/04/2023
[469] Ibid n59

stage. At present however, these shortcomings do act as a draw back and limit the potential effectiveness of the Act overall. For example, the Act would need to be adequately future-proofed, address the ever-changing nature of AI, consider the role of the typical end-user and the most at-risk fundamental rights in order to be truly effective in its approach.

Unlike the UK AI strategy however, the EU proposal is substantially fleshed out, and they do differ slightly in their proposed approaches. For example, the EU has taken a rather blanket strategy to regulating AI whereas the UK make clear via the National AI Strategy that they intend to employ a sector specific approach. This is somewhat similar to the technique used within the United States. Therefore, it is interesting to consider which approach, sector-specific or blanket, appears likely to work best.

4.3 Regulatory Strategies in Africa

Chapter Three of this thesis contains an examination of AI preparedness in the South African nation and this section highlighted some of the unique AI-based issues that South Africa faces in comparison to countries such as the UK and US. However, it is equally important to consider more broadly how Africa as a region is planning to cultivate and regulate AI. Most countries in Africa are yet to fully see the benefits of AI, benefits that are typically harnessed and enjoyed by countries within the global north. That is not to say however that Africa does not have a voice in the global conversation on AI, it most certainly does.

There are many countries within Africa that have already began to publish their own AI strategies including Mauritius[470] and Egypt,[471] as well as other nations that have not necessarily got a concrete strategy in place but are working towards one such as Kenya, Tunisia, Botswana, Rwanda, and Nigeria.[472] It is however worthwhile to consider the more broader scope AI initiatives which are being pursued at a regional level in Africa such as: the African Union (AU) Digital Transformation Strategy for Africa (2020-2030)[473] which will go on to form part of the AU's Agenda 2063,[474] the National AI Strategy that is being incentivised and pursued by both the AU Executive Council and AU High-Level Panel on Emerging

[470] Working Group on Artificial Intelligence, 'Mauritius Artificial Intelligence Strategy' (*ncb.govmu.org*, 2018) <https://ncb.govmu.org/ncb/strategicplans/MauritiusAIStrategy2018.pdf> accessed 01/04/2023
[471] The National Council of Artificial Intelligence, 'Egypt National Artificial Intelligence Strategy' (mcit.gov.eg, 2021) <https://mcit.gov.eg/Upcont/Documents/Publications_672021000_Egypt-National-AI-Strategy-English.pdf> accessed 01/04/2023
[472] G. Oloruntade, F. Omoniyi, 'Where is Africa in the global conversation on regulating AI?' (*techcabal.com*, 2023) < https://techcabal.com/2023/05/26/where-is-africa-in-the-global-conversation-on-regulating-ai/> accessed 28/05/2023
[473] African Union, 'The Digital Transformation Strategy for Africa (2020-2030)' (au.int, 2023) < https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf> accessed 01/04/2023
[474] African Union, 'Agenda 2063: The Africa We Want' (au.int, 2023) <https://au.int/en/agenda2063/overview> accessed 01/04/2023

Technologies,[475] and the work being undertaken on AI by the African Commission on Human and Peoples' Rights.[476] Therefore, this section of the thesis will consider these regional AI initiatives, draft strategies and proposed frameworks.

4.3.1 Agenda 2063

Agenda 2063 is good place to start as it is the AU's long-term strategy for transforming Africa into a 'key player in the global arena'.[477] The strategy is very wide-ranging in its aims, e.g., the agenda contains environmental goals, aims to promote peace and security, as well as developing the technical skills and education of its citizens to name just a few of its targets. AI has the capacity to feature widely within Agenda 2063, particularly with regards to education and skills development, transforming the economy and empowering its citizens.[478]

The Digital Transformation Strategy for Africa (2020-2030) (DTSA) goes some way in addressing AI as a part of this larger agenda.[479] Overall, the DTSA aims to create an inclusive digital society across Africa, with the primary focus being to establish Africa as a producer and not just a consumer of technology such as AI.[480] The DTSA aims to do this in a number of ways, but primarily by improving educational offerings to include tech-based curriculums, for example via a "massive e-skills development programme" to be delivered online to over 300 million Africans per year by 2025.[481] Therefore, the AU identifies its key route to creating a technologically innovative Africa is by creating and fostering a proactive learning environment, whereby citizens can gain the skills necessary to develop and work alongside emerging technologies such as AI.

Despite initiatives such as this being central to the agenda, further investment and development must take place in order to improve internet infrastructure across Africa to make access to programmes like the e-skills development scheme truly equal across the continent. This is something also addressed within the DTSA, with the aim being for everyone to have access to a reliable and stable internet connection by 2030.[482] If this aim is

[475] African Union Development Agency, 'The African Union Artificial Intelligence Continental Strategy for Africa' (nepad.org, 2022) < https://www.nepad.org/news/african-union-artificial-intelligence-continental-strategy-africa> accessed 01/04/2023
[476] African Commission on Human and Peoples' Rights, 'Declaration of Principles on Freedom of Expression and Access to Information in Africa' (achpr.au.int, 2019) < https://achpr.au.int/en/node/902#:~:text=The%20Declaration%20establishes%20or%20affirms,to%20express%20and%20disseminate%20information.> accessed 01/03/2023
[477] Ibid n474
[478] Ibid n474 'Goals and Priorities'
[479] Ibid n473
[480] Ibid n473
[481] Ibid n473
[482] Ibid n473

achieved, it could see Africa become one of the first regions to establish a wide-spread functional, digital, skills-based education for its citizens.

The DTSA also features the aim to have 99.9% of people within Africa be registered with a digital legal identity by 2030.[483] Digital ID systems are not unique to the African continent and are being invested in on a global scale. They typically involve biometric technologies that read facial features, fingerprints and iris scans, and help to prevent identity theft and help to make provision of government services more efficient.[484] These AI-powered systems are beneficial, but come with many ethical implications e.g. the potential for widespread discrimination; this is particularly the case with regards to facial recognition systems that usually form a part of a digital ID scheme.

The primary issue with regards to the facial recognition aspects of these technologies is that they are not developed in African countries, they are developed elsewhere (typically by tech companies in the global north) and are not trained on local facial data: this leads to misidentification and discrimination.[485] There are instances across Africa where this has proven to be the case already. Vumacam for example are a company that use a facial recognition system established in Denmark across their network of cameras in South Africa, where it has been proven that the technology frequently misreads and misidentifies African faces.[486] Similarly, in Uganda Huawei's facial recognition systems that were developed in China, were used in the 2020 election to find, and detain those supporting Bobi Wine.[487]

Therefore, whilst a continent-wide digital ID system might be beneficial in some respects, such a widespread system has the capacity to cause significant discrimination and infringement of human rights such as freedom of association, assembly, and privacy. The only way in which a system like this might work is if the AI itself was developed on the African continent or at least trained using a dataset that is representative of the communities it is to be deployed within. It is therefore logical that one of the key aims of the AU's digital strategy is to improve digital literacy across Africa, so that a generation of citizens with AI skills and knowledge will grow, and thus develop more reliable data sets and AI systems.

[483] Ibid n473
[484] Ibid n472
[485] Ibid n472
[486] Ibid n472
[487] K. Nkwanyana, 'China's AI deployment in Africa poses risks to security sovereignty' (aspistrategist.org, 2021) < https://www.aspistrategist.org.au/chinas-ai-deployment-in-africa-poses-risks-to-security-and-sovereignty/> accessed 01/04/2023

4.3.2 Work undertaken by the African Commission on Human and Peoples' Rights

Further to the above point on reducing instances of AI-based discrimination and bias within the African continent, it is crucial to look at the work being undertaken by the African Commission on Human and People's Rights. In 2019, the Commission published its Declaration of Principles on Freedom of Expression and Access to Information in Africa which states clearly that all states must ensure that any AI being developed and deployed across Africa must be aligned with international human rights.[488]

Further to this the Commission adopted a more detailed resolution in 2021 that asks the African Union to develop a regional regulatory framework that ensures that AI is used in a way that responds to the needs of the people of the continent.[489] This resolution is comprehensive in that it highlights many of the key issues Africa currently faces as a result of AI, such as the issues posed by deepfakes, misinformation and automated decision-making algorithms. Throughout, the resolution shows understanding of the potential for a variety of AI applications to impose upon all matter of human rights, which the Commission use as a basis for the need to undertake further study on the specific impacts AI might have upon the rights of its citizens.[490]

The resolution also emphasises the role that African states should play in the development of international AI policies and governance frameworks.[491] This point is key, as it ties in with the efforts promoted by the DTSA to establish Africa as a key player in the AI arena, but also reenforces the importance of international cooperation with regards to AI. It is essential that any AI technologies that are imported into the African continent are made applicable to African society, meaning that they are representative of the communities in which they are going to be deployed and are in line with the needs of the African people. This is acknowledged by the resolution, although no specific recommendation is made for how this might be achieved (aside from the mention of creating a governance framework that ensures this), it is an important and logical proposition.

4.3.3 Africa-EU Global Gateway

Building on this need for increased international cooperation and collaboration on AI regulation is the Africa-EU Global Gateway, a €150 billion investment from the EU to assist

[488] Ibid n476
[489] African Commission on Human and Peoples' Rights, 'Resolution on the need to undertake a Study on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa' - ACHPR/Res. 473 (EXT.OS/ XXXI) 2021
[490] Ibid
[491] Ibid

Africa in creating green and digital transformation.[492] It is specifically important with regards to AI as it is provides a strong basis for cooperation and collaboration to occur between the two regions. In particular this investment package will fund projects that improve digital infrastructure across Africa e.g., by installing fibre-optic cables for better broadband provision.[493]

As discussed within Chapter Three of this thesis with specific regard to South Africa, some scholars have suggested that it may well be the case that some African nations may choose to adopt regulatory strategies pioneered in other nations and regions, such as those already discussed in this thesis e.g., the EU. It is therefore possible that following collaboration via schemes such as the Global Gateway, the African region might favour an approach like that established within the EU AI Act.

However, it is worth bearing in mind the applicability of provisions, such as those within the EU AI Act, to Africa as a region. AI presents a number of unique challenges in Africa, and these challenges aren't necessarily shared with nations in the global north. Therefore, the applicability of legislative frameworks such as the EU AI Act, and other 'regulatory standards with extra-territorial application' to African nations may not be feasible.[494] Despite the importance of international cooperation and collaboration due to the global nature of AI, it is within the best interest of African jurisdictions and the African region to develop their own unique AI strategies that are specifically suited to the needs of the communities to which they will apply.

4.3.4 How do the strategies embed the principles of accountability, transparency, and non-discrimination?

The analysis in this section of the African region and its approach to AI regulation tells us a number of things regarding the importance of AI ethics in this region. The work being carried out by the African Commission on Human and Peoples' Rights via Resolution 473 is key to the incorporation of ethical AI principles in regulation within this region.[495] The resolution itself calls for the African Union to develop a regional regulatory framework that ensures that AI is used in a way that responds to the needs of the people of the continent, and most importantly addresses the potential human rights infringements that AI can result in.

[492] European Commission, 'EU-Africa: Global Gateway Investment Package' (commission.europa.eu, 2022) < https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/stronger-europe-world/global-gateway/eu-africa-global-gateway-investment-package_en#accelerating-the-digital-transition> accessed 01/04/2022
[493] Ibid
[494] Ibid n472
[495] Ibid n492

The solution to minimising and preventing these human rights infringements is typically by embedding principles such as accountability and transparency into AI systems, and ensuring individuals have opportunities to seek redress if and when harm does occur.

Further to this, the efforts made by the AU via Agenda 2063, to increase digital literacy and improve digital skills and education will go some way in ensuring that AI systems deployed in the African region embody the principle of non-discrimination. By the AU developing a workforce of digitally skilled individuals capable of creating all manner of AI systems, these systems will be more representative and better able to serve the needs of the communities in which they will be deployed, which is a key aim of the DTSA. Efforts currently underway in the African region therefore show both desire and promise to embed key ethical principles such as accountability, transparency, and non-discrimination within AI and AI governance.

4.3.5 Conclusions

The African region has an important voice in the global AI conversation and has a number of unique AI-related issues that aren't as prevalent in regions within the global north, such as the EU. This section has considered some of the AI initiatives and resolutions put forward within the African region in recent years, considering the strengths of these proposals as well as their limitations. Building on this examination, this thesis now goes on to consider efforts made by the UN to regulate AI.

4.4 Regulatory Strategies in the United Nations (UN)

There are numerous international bodies that have AI firmly placed at the forefront of their agendas, including the OECD, the World Economic Forum, and the UN. Out of these bodies, the UN is leading in regulatory developments in this space. The United Nations Activities on Artificial Intelligence report of 2022 noted that from 2021 to 2022 that 40 entities participated with the UN on AI related activities, there were 281 projects presented on AI and 84 new projects established, showing just how much effort the UN is dedicating to AI related issues.[496] This section will therefore examine several of these regulatory measures, with a view to considering their strengths, shortcomings, and possible room for improvement. This section is structured slightly differently than the previous sections in that it does not feature a specific section that considers how the UN measures embed key ethical principles within its regulatory approaches, this is because UN measures that specifically focus on embedding ethics within AI are analysed throughout this section.

[496] International Telecommunication Union (ITU), 'United nations Activities on Artificial Intelligence (AI)' (aiforgood.itu.int, 20220) < https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/> accessed 03/04/2023

4.4.1 UNESCO Recommendations on the Ethics of Artificial Intelligence

One of the most notable developments to come out of the UN regarding the regulation of AI is the Recommendations on the Ethics of Artificial Intelligence, released by UNESCO (United Nations Educational, Scientific and Cultural Organisation) in 2021 and adopted by its 193 member states.[497] This is the first real attempt to set the standard for ethical, safe, and reliable AI at a global level, and it encourages all member states to apply the recommendations in their own domestic setting. This is a responsibility that member states take on with the adoption of these recommendations, they must track and report their progress with regards to implementation of these measures, but UNESCO has vowed to support the 193 states in this.[498]

There are a number of key aims that the recommendations set out to achieve; firstly, to protect and respect human rights and fundamental freedoms from the impacts of AI, to promote equitable access to AI and share its benefits, and to truly embed ethics in all stages of the AI life cycle.[499] It would seem that the recommendations aim to do this primarily by ensuring better data protection, specifically by increasing transparency with regards to how a person's personal data is used. This is in line with some of the recommendations made earlier in this thesis (Chapter two), in that one of our primary lines of defence in combatting AI-related harms is via better data protection regulations; whether that means adopting new measures or amending old ones to make them more amenable to modern issues.

However, despite this document presenting a series of sound recommendations that go some way in achieving a united understanding of what ethical AI should look like, and ways we might achieve it, this document is not legally binding. This therefore means that whilst member states should do all they can to implement these recommendations into domestic policy, they are not legally required to. As per the recommendations, "Nothing in this Recommendation may be interpreted as replacing, altering or otherwise prejudicing States' obligations or rights under international law…".[500] These recommendations still remain a foundational document for establishing the key ethical principles for AI, but they will not have the same legal weighting as the EU AI Act for example (once it is in force). Although, the

[497] UNESCO, 'Recommendations on the Ethics of Artificial Intelligence' (unesdoc.unesco.org, 2022) <https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi> accessed 01/04/2023
[498] UNESCO, 'UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence' (*unesco.org*, 2021) <https://www.unesco.org/en/articles/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence> accessed 01/04/2023
[499] Ibid n497
[500] Ibid n947

recommendations are available to a much wider global audience than the EU regulations will be.

The recommendations contain a section that specifically highlights policy areas of concern and how member states can implement effective measures in these areas in order to meet the values set out within the document. Some of the notable policy areas of concern include data policy, environment and ecosystems, culture, education and research, economy and labour, and health and social well-being.[501] In this section of the recommendations, UNESCO promotes the beneficial use of AI across these sectors, giving examples of ways in which nations might utilise AI in museums and galleries for example or to prevent global warming. This means that whilst the recommendations are concerned with ensuring the development of safe and trustworthy AI, the outlook of the document on the future of AI is generally rather positive, focussing on its benefits as opposed to the harm, it can cause.

These policy recommendations are supplemented by the recommendation that governments across the world should establish and encourage use of ethical impact assessments; when using AI across sectors, organisations should be conducting these assessments to ensure that any AI use is responsible and ethical. The recommendations do not give a comprehensive example of what such a risk assessment might look like, other than that these impact assessments should establish what potential impact the application might have upon human rights and fundamental freedoms. This is in line with the recommendations made by this thesis, and Chapter Six provides a proposal for a rights-based impact assessment that aligns with the ethical recommendations made here by the UN.

In a similar vein to the general recommendations adopted in 2021, the UN System Chief Executives Board for Coordination endorsed the Principles for the Ethical Use of Artificial Intelligence in the United Nations System.[502] This is a document that is specific to the use of AI systems in the UN, and it guides the use of such systems to ensure that they meet the recommended ethical principles endorsed by UNESCO. This is just one specific example of the UN essentially following their own recommendations in-house, which shows commitment to their endorsement and that the recommendations are in fact reasonable and workable. This particular document could even act as a template for organisations wishing to implement the UN's ethical recommendations.

---

[501] Ibid n497

[502] Chief Executives Board for Coordination, High-Level Committee on Programmes, Inter-Agency Working Group on Artificial Intelligence, 'Principles for the Ethical Use of Artificial Intelligence in the United Nations System' (unsceb.org, 2022) < https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf > accessed 03/04/2023

Therefore, UNESCO's ethical recommendations seem to provide a solid starting point for meaningful AI regulation. Whilst the recommendations are not legally enforceable in the same way that other legislation might be, it is the first instrument that sets the standard for ethical AI on a global level, therefore making it a valuable document. There are, however, other UN measures to combat the issues posed by AI that are necessary to examine.

4.4.2 Other AI initiatives

It is worth considering the future efforts that look to be made by the United Nations Interregional Crime and Justice Research Institute (UNICRI) with regards to AI and robotics for crime prevention, criminal justice, law enforcement and national security.[503] The centre recognises the benefits of AI within crime prevention, e.g.in monitoring criminal networks and predicting the commission of future criminal acts etc. However, the centre also acknowledges the real ethical and legal issues associated with using AI in these ways; as discussed earlier in this thesis there are significant transparency issues, and room for mass discrimination when using AI in this capacity.

The centre is therefore dedicating efforts to establish how to use AI in the criminal justice system in a responsible way. They are doing this by collaborating with industry specialists such as standards development bodies like International Telecommunications Union (ITU), the International Criminal Police Organisation (INTERPOL) and other international organisations such as the World Economic Forum (WEF).[504] These are some of the key stakeholders that UNICRI are targeting with their work, alongside policymakers within UN member states. Some of the activities that the centre is endorsing include the creation of risk assessment frameworks for the use of AI within the criminal justice system, and training key stakeholders as to the benefits and risks of using AI in this capacity.

There is therefore a trend that can be spotted here; UN agencies are actively encouraging the use of AI across sectors within its member states, focussing on the benefits of the technology but also proactively educating stakeholders on the ethical risks simultaneously. This is in line with the recommendations made within this thesis in that we must carefully balance the encouragement of innovation with the need to ensure the develop safe AI.

One of the key recommendations and outcomes from the ongoing conversation within the UN on AI is to appreciate the usefulness of risk assessments to evaluate the impacts of AI. The UN is endorsing the use of risk assessments that assess the impact of AI upon human rights and fundamental freedoms across all sectors. This could be compared to some extent

[503] UNICRI, 'Centre on Artificial Intelligence and Robotics' (unicri.it, 2023) <https://unicri.it/topics/ai_robotics/centre> accessed 03/04/2023
[504] Ibid

to the efforts made within the EU AI Act to categorise AI applications based upon their risk of causing harm, however, a comprehensive rights-based impact assessment, similar to the one endorsed by the UN seems more functional.

In a similar vein to the other efforts pursued by the UN discussed above, the UN established AI for Good, "…a year-round digital platform where AI innovators and problem owners learn, build and connect to identify practical AI solutions to advance the UN SDGs."[505] It is the leading global UN-lead platform on AI. This particular programme is organised by ITU as well as 40 UN agencies and aims to find practical solutions to help achieve the UN Sustainable Development Goals.

AI for Good holds a summit most years in which stakeholders from all over the world can convene to discuss AI related issues and work towards achieving practical solutions to these problems. Platforms such as this are incredibly useful as they allow for necessary conversation to take place between those most invested in the development of AI and allow for collaboration to take place on an international scale, where it might not have done previously.

Further to this there are UN conventions such as the Convention on Certain Conventional Weapons (CCW) or the Inhumane Weapons Convention which gives consideration to human control as it relates to AI technologies.[506] In particular, this convention considers human responsibility for the decisions on the use of weapons, as accountability cannot be transferred to the system itself, and any weapons systems that are based on emerging technologies such as AI should always comply with international humanitarian laws. This is a particularly promising convention that provides a good entry point for AI regulation in that whilst specifically dealing with autonomous weaponry, it considers the human-control aspect of AI, and how accountability for certain AI-driven systems must always come back to a responsible human.

4.4.3 Conclusions

The UN is arguably the leading international organisation in terms of AI regulation. The UN have done important groundwork in establishing a global standard for ethical AI that has been adopted by 193 member states around the world. Whilst there is some way to go in ensuring these member states fully implement these recommendations to the best of their

[505] AI for Good, 'About' (aiforgood.itu.int, 2023) < https://aiforgood.itu.int/about-ai-for-good/> accessed 03/04/2023
[506] United Nations Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects as amended on 21 December 2001

ability, this is a crucial step in the road towards better AI regulation. There are a number of other AI initiatives being pursued by the UN that focus on AI applications in specific sectors, such as the criminal justice system as considered above. Therefore, the UN has asserted itself as a leading voice in the conversation on AI regulation.

4.5 Conclusions

This Chapter has examined approaches to AI regulation from both regional and international perspectives, considering how the EU intends to tackle AI, how the African region are approaching AI, as well as some of the measures endorsed by the UN such as the Recommendations on the Ethics of AI.

Each of these regions and international bodies seem to prefer to take slightly different approaches to dealing with AI. For example, the EU AI Act is a blanket-style measure that categorises AI based upon risk-factor, whilst one prominent component of the African regional approach to AI is to focus on promoting digital education as a way to tackle AI associated risks. This can be compared to the UN approach, which is the widest reaching in scope, and endorses the use of risk-assessments that specifically consider the impact of AI upon human rights and fundamental freedoms.

Each of these approaches have their strengths; in the case of the African continent, focussing on improving AI-literacy has the capacity to ensure that home-grown AI is much more representative of the communities it is applied to, and thus more ethical than perhaps an imported AI developed in the global north. Whilst within the EU, categorising AI based upon risk factor is a logical option. And the UN's foundational ethical recommendations promote the use of AI across sectors so that the global community can equally share its benefits.

Yet, some common shortcomings or potential issues can be identified, e.g., in the case of Africa, it is necessary for general internet infrastructure to be improved across the continent in order for there to be equal access to the AI/digital curriculum intended to benefit all citizens. The EU's risk-based approach has some flaws in terms of its categorisation of applications and the wording used within the Act being rather narrow and misleading. In addition, the UN's ethical recommendations are not legally binding, and so whilst they are a good starting point it will be down to the individual member states to do the important work and implement these principles in future regulation.

Chapters Three and Four consider a variety of AI strategies, frameworks and initiatives developed by several nations, regions and international organisations and consider their strengths and weaknesses, with some recommendations made for amendments to these.

Therefore, these Chapters address the second and fourth research questions set out within this thesis, these are: How well equipped are current legal instruments, proposed legal instruments, and strategies in dealing with the issues posed by AI? And what realistic and workable recommendations can be made to improve the current state of AI regulation?

4.5.1 Recommendations to regional bodies on the regulation of AI

There are a number of recommendations made to regional bodies on the regulation of AI, which are as follows:

In a similar vein to the recommendations made to nations and states in the previous chapter, regions should consider carefully what their primary interests are in pursuing AI regulation, and what unique barriers states might face in implementing AI with their regions, which will enable them to create a bespoke and fitting regulatory regime that serves that specific region. For example, as demonstrated in this analysis of the African region, a primary goal needs to be the improvement of digital infrastructure across the region to enable African populations to equally benefit from AI. This is in comparison to the EU which by and large has this infrastructure in place, therefore meaning that this barrier is not applicable here. As a result, EU-style regulations mightn't be all that suitable in the African region.

Pursuing opportunities for international collaboration between regions is also desirable and recommended by this thesis. The Africa-EU Global Gateway that is currently underway in the two regions is just one example of how regions can work together to achieve policy goals. This specific scheme is multifaceted, with one aspect being to focus on the development of digital infrastructure and AI adoption within the African region. Collaborative agreements such as this are beneficial as they enable regions to lend expertise to one another and benefit from one another in areas of need, such as AI regulation.

One unique challenge for regional bodies is going to be creating regulatory regimes that are applicable across regions, this may be difficult to orchestrate due to the diverse needs of member states within a given region. As a result, careful consideration must be given by regional bodies attempting to regulate AI in that their regimes must, to the best of their ability, be reflective of and serve all of their member states.

4.5.2 Recommendations to the United Nations on the regulation of AI

As with the recommendations made to regional bodies in the previous section, it is also recommended that the UN gives careful consideration to the measures they create and their applicability to the vast number of UN member states. It will be easier for some states to implement guidelines than it will be for others that have more complex barriers to AI implementation. However, drawing upon initiatives such as the UNESCO Recommendations

on the Ethics of Artificial Intelligence, whilst member states should do all they can to implement these recommendations into domestic policy, they are not legally required to. This causes some concern regarding the overall effectiveness of the recommendations, but nonetheless does not impose unreasonable requirements on states at different levels of AI preparedness.

It is also recommended that the UN continues its pursuit of calling for governments across the world to establish ethical impact assessments. The UN asks that when using AI across sectors, organisations should be conducting these assessments to ensure that any AI use is responsible and ethical. This is in line with the detailed recommendations made in Chapter Six of this thesis. Although, the UN recommendations do not give detailed example of what such a risk assessment might look like, they state that these impact assessments should establish what potential impact the application might have upon human rights and fundamental freedoms. It would be useful for the UN to perhaps establish more detailed templates of these impact assessments that could be used by governments and organisations around the world. An example of what this might look like is proposed within Chapter Six of this thesis.

Chapter Five

Regulating AI: The ideal regulatory response?

5.1 What does the ideal regulatory response look like?

Creating a new regulatory framework within a rapidly changing field is no easy venture, as displayed in previous chapters. Particular care must be paid to the type of governance measures selected and relied upon for a number of reasons, so the question is: where do we begin?

Drawing upon the approaches taken by different countries within the same field is a good place to start; as international issues ultimately benefit from transnational, aligned approaches.[507] We must also be conscious that within any accepted regulatory approach to the uses of artificial intelligence, that we ensure that key concepts are embedded within its essence: namely transparency, accountability and responsibility must be at the heart of the response.[508] Another equally important and foundational factor that must be given heed to is the sheer interdisciplinary and overlapping nature of AI; how can we effectively govern a technology with such wide-spread impact and effects?[509] Piecing together these separate elements to create an overall effective and working governance system for AI will be a complex task, but one that is achievable given sufficient consideration.

One thing is for certain: regulation is necessary in order to govern the development and various uses of AI, whether this be via centralised regulation or by other means.[510] Arguably, this is something that lawmakers have so far tried to postpone or delay considering in detail, likely due to the intricate and sensitive nature of the task at hand.[511] Although more recently, there appears to be a growing global appetite for considering how we can work together to ensure the safe and secure use of AI; for example, as demonstrated previous chapters, the new Biden administration have made clear their intentions to 'reengage with the world', which is particularly encouraging seeing as the US is one of the lead players in the AI

---

[507] O.J. Erdelyi, J. Goldsmith, 'Regulating Artificial Intelligence: Proposal for a Global Solution' (2018) *AIES '18: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society,* February 2–3, 2018, New Orleans, LA, USA, 95-101, 96

[508] W. Hoffmann-Riem, 'Artificial Intelligence as a Challenge for Law and Regulation' in T. Wischmeyer and T. Rademacher (eds), *Regulating Artificial Intelligence* (Springer, 2020)

[509] C. Cath, 'Governing artificial intelligence: ethical, legal and technical opportunities and challenges' (2018), *Philosophical Transactions, Royal Society Publishing* <https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2018.0080> accessed 01/06/2021

[510] Ibid n508

[511] Hon. Justice Michael Kirby, 'The fundamental problem of regulating technology' (2009) *The Indian Journal of Law and Technology,* 5 1-32

sphere.[512] Similarly, China has committed to become a 'driving force' in pioneering ethical norms and improving the standards for AI in their 'New Generation Artificial Intelligence Development Plan' published in 2017.[513] And likewise, the UK government published in March 2021 its intentions to "make the UK a global centre for development and commercialisation" by adopting responsible AI in line with its new AI strategy.[514]

Despite the shortcomings demonstrated in these approaches, one positive commonality that appears throughout these international commitments to improve our 'digital future' is the proposal that we must make AI more trustworthy.[515] This is very high-level, but not without warrant. Agreement on the need to improve the trustworthiness of AI suggests that prior to the development of any robust legal regulatory response, we need to simultaneously ensure the creation of technologies that respect both fundamental rights, such as the right to protection from discrimination, and key ethical principles.[516]

To reiterate, when ethical principles are mentioned, this is in reference to principles identified by nations and organisations across the globe, with the most commonly agreed upon ethical principles for AI being transparency, accountability, and non-discrimination. Some of the most notable ethical guidelines that identify these key principles include the EU's High-Level Expert Group on Artificial Intelligence (HLEG) Ethics Guidelines for Trustworthy AI, which specifically highlights technical robustness and safety, as well as transparency, accountability and non-discrimination as key ethical principles.[517] The SHERPA guidelines on development and use of ethical AI expanded upon the work of HLEG and add to this list the benefit of embedding ethics by design into AI systems, an approach that is supported by this thesis.[518]

---

[512] J.P. Meltzer, C.F. Kerry, 'Strengthening international cooperation on artificial intelligence' (*Brookings, 17 Feb 2021*) <https://www.brookings.edu/research/strengthening-international-cooperation-on-artificial-intelligence/> accessed 01/03/2021

[513] H. Roberts, J. Cowls, J. Morley, M. Taddeo, V. Wang, L. Floridi, 'The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation' (2021) *AI & Society,* 36, 59-77

[514] Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, Office for Artificial Intelligence, 'New strategy to unleash the transformational power of Artificial Intelligence' (*GOV.UK,* 12 March 2021) https://www.gov.uk/government/news/new-strategy-to-unleash-the-transformational-power-of-artificial-intelligence accessed 15/07/2021

[515] European Commission, 'High-level expert group on artificial intelligence' (*digital-strategy.ec.europa.eu,* 23 June 2021) <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai> accessed 02/07/2021

[516] European Commission, 'Member States and Commission to work together to boost artificial intelligence "made in Europe"' (*digital-strategy.ec.europa.eu,* 7 December 2018) <https://ec.europa.eu/commission/presscorner/detail/en/IP_18_6689> accessed 02/07/2021

[517] Ibid n468

[518] SHERPA, 'Guidelines for the Ethical Use of AI and Big Data Systems', and 'Guidelines for the Ethical Development of AI ad Big Data Systems: An Ethics by Design approach' (*project-sherpa.eu*, 2020) < https://www.project-sherpa.eu/guidelines/> accessed 04/04/2023

In addition, there are a plethora of other ethical guidelines and recommendations that confirm the most important ethical principles we must ensure we respect as we develop and use AI. A number of standards development organisations have also contributed to this list of ethical principles for AI, including the British Standards Institute (BSI) via BS 8611 on the ethical design and application of robots and robotic systems,[519] and the Institute of Electrical and Electronics Engineers (IEEE) via their Global Initiative on Ethics of Autonomous and Intelligent Systems.[520]

Therefore, by utilising the influence that international standards development organisations have upon industry, particularly in the digital space, we can begin to piece together a multi-faceted approach to regulating AI.

A similar approach has already been taken for various types of emerging technologies, such as for the improvement of consumer IoT (Internet of Things) cybersecurity. Here, the UK government have worked in conjunction with the European Telecommunications Standards Institute (ETSI) to develop industry standard EN 303 645 which creates a cybersecurity baseline for consumer connected products.[521] This particular standard is embedded within the UK's Product Safety and Security Bill, in which a common security baseline (including specific device security requirements) is mandated by law for consumer IoT products sold in the UK.[522]

This combined approach to regulating and improving the security of consumer IoT devices was deemed to be most effective due to the number of industry stakeholders implementing the standard. I had personal experience working on this project, and during a number of consultancy exercises we were able to establish that it would be much easier to mandate certain requirements by law in the UK if we first created a standard that manufacturers would be more likely, and more incentivised to take up. As a result, it is proposed that this is an equally desirable approach to take for AI; a possible combination of industry standards that

[519] British Standards Institute, 'BS 8611 Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems'(standardsdevelopment.bsigroup.com, 2016) < https://standardsdevelopment.bsigroup.com/projects/2022-00279#/section> accessed 04/04/2023
[520] IEEE, 'The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems' (standards.ieee.org, 2017) < https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_general_principles_v2.pdf> accessed 04/04/2023
[521] European Telecommunications Standards Institute, 'Consumer IoT Security' (*ETSI.org*) <https://www.etsi.org/technologies/consumer-iot-security> accessed 02/07/2021
[522] Department for Digital, Culture, Media and Sport, 'Government response to the call for views on consumer connected product cyber security legislation' (GOV.UK, 21 April 2021) <https://www.gov.uk/government/publications/regulating-consumer-smart-product-cyber-security-government-response/government-response-to-the-call-for-views-on-consumer-connected-product-cyber-security-legislation> accessed 05/05/2021

target the tech and centralised governance in the form of legislation seem to be a successful pairing.

Therefore, this chapter will consider the options available for regulating AI, and how we can begin to build a functioning regulatory ecosystem, answering research question three; what key regulatory principles are valuable and should be included in an ideal AI governance framework? In particular, this chapter draws upon current approaches taken to regulate modern technology in related areas, such as the aforementioned approach taken to improve the security of consumer IoT (Internet of Things) devices,[523] and considers how we can potentially use similar principles and paths in order to effectively regulate AI.

In order to establish the 'ideal' framework for harmonisation for AI, a variety of methods, concepts, and models will be examined in this chapter, such as: utilising both primary and secondary legislation in the most effective ways, development and use of industry standards, the importance of international harmonisation and alignment, and the principles that underpin AI.

5.2 The Five Paradoxes

The emergence of more sophisticated and intelligent technologies during the past couple of decades has served as a catalyst for change, in a rather static and traditional regulatory environment. Regardless of the technology in question, the same issues remain; to effectively regulate modern tech we must find a way to continue to promote innovation, protect the interest of consumers, and also protect society from the somewhat unpredictable consequences arising out of the use of various modern technologies.[524]

Therefore, it seems pertinent to begin this chapter by first considering the 'Five Paradoxes' of regulating technologies as described by the Honourable Justice Michael Kirby.[525] These paradoxes, or 'curiosities' as he otherwise calls them, came to fruition following debate at the TELOS conference in April 2007, a conference hosted by the Centre for Technology, Ethics, and Law in Society at King's College School of Law, London. Here the topic of regulating technologies was heavily debated, and as a result five paradoxes became apparent.

These observations are particularly insightful and useful to consider in light of the issue at hand, i.e., the regulation of AI. They perhaps suggest why it appears as though regulating

[523] ibid
[524] W. D. Eggers, M. Turley, P. Kamleshkumar Kishnani, 'The future of regulation: Principles for regulating emerging technologies' (*Deloitte Insights,* 2018) <https://www2.deloitte.com/us/en/insights/industry/public-sector/future-of-regulation/regulating-emerging-technology.html> accessed 15/07/2021
[525] Ibid n516 p. 11

AI, and other modern technologies, has presented itself as a contentious issue in recent years. Work such as that undertaken by Justice Kirby is especially useful in the context of this chapter, as it presents some of the common obstacles and challenges that come with creating a functioning and effective, modern regulatory system. The five paradoxes listed in this work are as follows: lack of expertise, too much/too little law, the issues presented by copyright laws, the impact that technology has on democracy and that the topic has often been neglected within legal circles.[526]

Whilst these paradoxes are useful and indicative of some of the problems that will certainly arise when considering the regulation of technologies such as AI, a number of these challenges have been somewhat lessened in recent years, although to some extent they are still present.

Lack of expertise

Starting with the first of the paradoxes, the lack of expertise in the subject of regulating technologies, it is worthwhile noting for the purposes of this thesis that this has significantly changed within the last decade.[527] As our understanding of new and emerging technologies has developed, the number of academics, researchers and scholars in the area has grown too. A simple web search indicates the growing number of universities across the world offering courses and modules specialising in the study of artificial intelligence and its relationship with law and regulation. This is most definitely promising and signals a rise in the number of scholars and practitioners working in the field, who are capable of delivering this teaching to students.

Despite this change however, it still is true that it is much easier to find a specialist or expert in tax law than it would be to find one for the regulation of artificial intelligence.[528] It is therefore worth considering the impact that this might have upon the development of regulation in this area; if there are a considerably smaller number of experts in this field than in other areas of law, then there is a higher likelihood for the strength and type of the regulations chosen here to be markedly shaped by the opinions and thoughts of the few

[526] Ibid n516 p. 11-23
[527] See for example The Alan Turing Institute: https://www.turing.ac.uk/. This is one of the leading institutes for data science artificial intelligence and is based within the UK. The institute has a number of research clusters that deal with a plethora of issues related to AI, from data ethics and machine learning to privacy and theoretical mathematics. The institute relies on the work of excellent, and world-leading researchers, and so with the growth of institutes like this, the number of qualified scholars and academics in this field will continue to grow.
[528] Ibid n516 p. 12

experts that do exist and work on the task. Whilst this is unavoidable, it is most certainly worth bearing in mind.

Striking the balance between too much or too little law

The second of the paradoxes is arguably the most crucial of all; striking the balance between laws that do not go far enough (or even exist in the first place) and laws that go too far and inhibit worthwhile technological development.[529] This chapter (and more widely, this thesis) considers how to strike this balance at length, and via the suggestions made in this chapter, the aim is to create a regulatory framework for AI that holds this balance at its centre.

The issue we face is that by not regulating at all, or by imposing very minimal regulations and requirements upon scientists and those within industry, we would essentially be letting the moral and ethical beliefs of these individuals govern the AI industry. This is not good enough, and due to the nature of the risks that can arise from the use of AI (risks to physical safety, societal impacts etc) more will need to be done.

Alternatively, taking an overzealous approach to regulating AI can have a chilling effect on the actual development of the technology and is completely counterintuitive; the technology in question might ultimately prove extraordinarily beneficial to us. Therefore, in order to adequately regulate and govern this particular area, a 'consistent and positive narrative' is necessary, however, we realistically do not have this consistent narrative yet.[530] Thus, more needs to be done to promote the use of responsible AI whilst ensuring that we have the correct and proportionate safeguards in place to guarantee its appropriateness for use.

Intellectual property rights

For the purpose of this thesis, there appear to be two distinct issues with regards to the relationship that exists between intellectual property law and artificial intelligence; access to algorithms that are classed as trade secrets, and censorship. By instilling regulatory measures within the algorithm or code of the AI itself, it is suggested that we might begin to excessively censor what information or materials are available to us, inhibiting our right to freedom of expression. The most common example as used by Justice Kirby, is one of inserting a filter that stops underaged users from accessing 'harmful' material.[531] This initially makes sense, however, where is the line drawn with regards to what information can be classed as 'harmful'?

[529] Ibid n516 p. 15
[530] Ibid n516 p. 8
[531] Ibid n516 p. 17

In a similar vein, the issue of using closed-source algorithms for the purpose of imposing punitive measures on individuals, should be noted. In the case of Loomis v Wisconsin[532], the lack of transparency with regard to the closed-source risk assessment software used to determine the defendants recidivism rate, caused a claim to be raised regarding the algorithms validity and accuracy in carrying out such a task. Access to the algorithm itself was refused by its owner, as it was classed as a trade secret. This presents a challenge going forward; in order to properly regulate AI, it is likely that for enforcement purposes, access to closed-source algorithms may be necessary. The answer to this issue may lie within the new regulatory framework itself, or in the amendment of existing legislation that governs this area.

Impact on democracy

This particular issue is one that has been the subject of hot debate in recent years, most notably since the Cambridge Analytica scandal in 2018/19. One question that we must consider when regulating AI is how does this modern technology interact with our democratic values? A point raised again by Justice Kirby that is worth considering is that we must think hard about how we can make large multinational industry players like Apple or Google subject to the democratic values of the states in which they sell their products?[533]

The topic has been neglected within legal circles

The final of the five paradoxes, but an interesting one to consider. Earlier on, the point was made that there appeared to be a lack of expertise in the field when compared to other legal specialisms. The real reasoning behind Justice Kirby's inclusion of this final paradox is to encourage all those who have any interest in the rule of law to engage in the study of this particular field.[534] The nature and scale of the impact that the evolution of modern technology, specifically that of AI, is a topic worthy of considerable attention and resources.

The five paradoxes as discussed here set out the principal issues and challenges that we must consider when embarking on the task of regulating any modern technology, including AI. As this chapter considers how we can begin to build an effective regulatory system for AI, these paradoxes can act as a check list of sorts; they should be kept in mind and considered alongside any meaningful regulatory proposal.

[532] 881 N.W.2d 749 (2016)
[533] Ibid n516 p. 21
[534] Ibid n516 p. 22

This chapter will now go on to consider what the ideal regulatory response to AI looks like; this includes identifying the most suitable governance measures and making suggestions for hybrid approaches where necessary.

5.3 Transparency: the key ethical principle

Before we begin to consider the actual methods that we can use to legally regulate AI, we need to first consider what exactly it is that we would like to achieve by introducing legislation or developing industry standards for example. The simple answer to this is that we want to make AI safer; we want to protect citizens from harm but continue to promote innovation and growth within emerging tech industries. However, in order to achieve this rather high-level outcome, we need to consider specifically how we can go about ensuring one of the key principles underpinning AI is embedded within regulatory efforts; this principle being transparency.

The very nature of AI makes it infinitely more difficult to regulate; for example, in the physical world we can use precise statistics, models and decade-long experiments to predict the likelihood of a physical disaster occurring in a particular place at a particular time, and we can use this knowledge to inform proportionate and reasoned responses. However, when it comes to artificial intelligence, we are not so fortunate. There are a number of variables and factors that affect the functioning of even the simplest of artificially intelligent systems, and we must be aware of these intimate details prior to developing any robust regulatory response.

Therefore, the following rings true:

> "By far the greatest danger of Artificial Intelligence is that people conclude too early that they understand it."[535]

Yudkowsky captures within this quote my primary concern when it comes to any attempt to regulate AI. We are arrogant if we believe that we truly understand AI, and that we are capable of completely regulating it as things currently stand. To truly protect the world from the legitimate global risks that AI presents, we must analyse the core principles that underpin artificial intelligence and identify those that are causing issue at present. Three key principles have been referenced throughout this thesis, namely accountability, transparency, and non-discrimination. It is arguable however the most important principle of the three is

[535] E. Yudkowsky, 'Artificial Intelligence as a Positive and Negative Factor in Global Risk' in N. Bostrom and M. M. Ćirković (eds), *Global Catastrophic Risks* (Oxford University Press, 2008)

transparency; by pursuing transparency within AI, it becomes easier to ensure accountability and minimise discrimination.

Therefore, it is proposed that we begin the process of building the 'ideal' regulatory system for AI by considering the principles listed above: transparency in some detail. By understanding the role that transparency plays within AI, and the role that it can play within AI governance in the future, we can therefore set the foundations for a functioning regulatory system.

There is some contention however regarding how we actually decide which principles are worthy of note and embedding within AI; the issue here lies within personal competing conceptions of value, which is a valid concern. [536] Although, to this point it is contended that the principles proposed for consideration and encoding within this thesis (namely accountability, transparency, and non-discrimination) are commonly agreed upon as being the foundational and fundamental principles in need of immediate consideration.[537]

By considering transparency in some detail, it is proposed that this will allow us to better design a functional and effective regulatory response (whether that be via legislation, developing industry standards, or a combination of approaches) with awareness of this key principle embedded at its core.

5.3.1 Transparency

Transparency is a concept at the core of most ethical discussions on AI; in fact, transparency is the single most common principle cited within ethical AI guidelines (a study from 2019 found that the principle of transparency was listed as key principle at least 84 times).[538] However, as with the term AI itself, we must consider more closely what we mean by 'transparency' in this particular context.

Our understanding of the term transparency can very much change depending upon the context within which we consider it, much like AI. And with regards to AI, transparency can also mean a number of things, for example we might be referring to algorithmic transparency in which we would look more closely at how exactly an algorithm reached a particular decision. Or we might want to know to what degree we able to view the various ways

[536] I. Gabriel, 'Artificial Intelligence, Values and Alignment' (2020) *Minds and Machines,* 30, 411-437
[537] U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl, J. D. Weisz, 'Expanding Explainability: Towards Social Transparency in AI systems' (CHI Conference on Human Factors in Computing Systems, May 2021)
[538] S. Larsson, F. Heintz, 'Transparency in artificial intelligence' (2020) *Internet Policy Review,* 9(2) referencing a study included within A. Jobin, M. Ienca, E. Vayena, 'The global landscape of AI ethics' (2019) *Nature Machine Intelligence,* 1 389-399

artificially intelligent systems are being used around us, and what degree of accountability there is for any potential shortcomings that may negatively affect us?

As such, and for the purposes of this work, a comprehensive, multidisciplinary understanding of the term 'transparency' is used. Here we will consider the principle of transparency both from a technical and algorithmic perspective and from a governance standpoint, with the aim being to bridge the existing gap between the concepts of accountability and transparency present within the literature.[539] We therefore understand transparency to mean openness, knowledge and understanding, and the removal or limitation of the typical 'black box' phenomenon that usually accompanies AI.[540] Transparency can also act as a tool that allows us to assess a systems reliability and its behaviour, which is relevant for any sophisticated AI.[541]

Therefore, the tie between the concept of transparency and trust is firm; the European Commission have made this very clear in their White Paper on the topic of the future of AI within the EU.[542] If we wish to reap the benefits that AI presents us with whether that be via advances in healthcare, business development or for the greater public interest,[543] then we must ensure that artificially intelligent systems being used are fair and trustworthy, and fairness is achieved by ensuring transparency.[544]

Yet the level of transparency that we desire and require of any given AI system will likely change depending upon the uses of that system itself. For example, there is a clear and growing concern regarding the lack of transparency within most algorithmic decision-making processes.[545] If we have an AI being used to filter the films that we're most likely to enjoy on our favourite streaming service, the need and desire for transparency regarding the way in which the system presented us with those choices is minimal; we may be intrigued to know how that choice was made, but the stakes and resulting impacts for us are not that high. However, if an AI is being used to decide whether a person is eligible for a bank loan or admission to a university, the stakes (and therefore risk of bias and discrimination) are much

---

[539] Ibid
[540] M. Ananny, K. Crawford, 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability' (2016) *New Media and Society* <https://journals.sagepub.com/doi/10.1177/1461444816676645> accessed 10/06/2021
[541] A. Theodorou, R. H. Wortham, J. J. Bryson, 'Designing and implementing transparency for real time inspection of autonomous robots' (2016) *Connection Science* 29(3) 230-241
[542] Commission, 'White Paper on Artificial Intelligence – A European approach to excellence and trust' COM (2020) 65 final
[543] Ibid n514
[544] B. Lepri, N. Oliver, E. Letouze, A. Pentland, P. Vinck, 'Fair, Transparent and Accountable Algorithmic Decision-making Processes' (2018) *Philosophy & Technology* 31, 611-627
[545] See Chapter Two and J. Graham, 'Risk of discrimination in AI systems: Evaluating the effectiveness of current legal safeguards in tackling algorithmic discrimination' in A. Lui, N. Ryder (eds) FinTech, Artificial Intelligence and the Law: Regulation and Crime Prevention (Routledge, 2021)

higher, meaning that there is also a more legitimate need and desire for transparency within that process.

Thus, the higher the risk and the higher the potential impact that using an artificially intelligent system may have upon an individual, a higher degree of transparency is required. As a result, in order for us to meaningfully incorporate the concept of transparency within any effective governance structure for AI, it is reasonable to suggest that a risk-based approach which allows for the level of transparency required to be based upon the degree of potential impact likely to be caused by an artificially intelligent system, is a sensible recommendation. Further consideration as to what this risk-based approach might look like will be discussed in detail further on in this chapter.

5.3.1.1 Introducing 'transparency by design'

It is therefore clear that within any effective regulatory response, transparency must be considered as a cornerstone, but how can we best incorporate this notion within our proposed governance structure? Legislative measures could go some way in ensuring that we have transparent AI, however, more direct impact could be made by embedding transparency by design principles within AI.[546] The concept of embedding a principle such as transparency within the design of a product or system has proven a popular choice; for example, with the introduction of GDPR (General Data Protection Regulations), organisations and public bodies had to implement Privacy by Design principles within their technology.[547] This meant that privacy principles had to be embedded within systems and processes by design, and not merely considered as an afterthought or only when issues arose.

Similarly, the work undertaken by the UK government to secure consumer IoT devices employs a similar methodology[548]; here the aim however is to embed cybersecurity principles by design within the technology itself as opposed to ethical principles like we are dealing with in this thesis.[549] Nonetheless, the approach adopted by the UK government regarding IoT devices could be adopted and adapted for use in AI regulation, to embed ethical principles such as transparency within the technology itself. Ultimately, these devices

[546] H. Felzmann, E. Fosch-Villaronga, C. Lutz, A. Tamo-Larrieux, 'Towards Transparency by Design for Artificial Intelligence' (2020) *Science and Engineering Ethics* 26, 3333-3361
[547] Privacy Policies, 'Implementing Privacy by Design' (privacypolicies.com, 5 January 2021) <https://www.privacypolicies.com/blog/privacy-by-design/#What_Is_Privacy_By_Design> accessed 05/08/2021
[548] Ibid n527
[549] Please see here an overview of the work undertaken by the UK's Department for Digital, Culture, Media and Sport, Secure by Design team. <https://www.gov.uk/government/collections/secure-by-design> accessed 19/05/2021

including artificially intelligent systems, are products and as such incorporating features within their design that will make them as safe as possible for end users is essential. This notion aligns with rules three and four of the 'Principles for Designers, Builders and Users of Robots'; notably that such 'intelligent' devices should be designed using processes that assure their safety and security, and that such devices should not be designed in a way that could exploit vulnerable users (e.g., they should be as transparent as possible).[550]

By treating artificially intelligent systems in this way, as products that are required to be both reliable and safe for end users, following similar approaches taken for both data protection purposes and for consumer IoT, embedding transparency as a principle by design into an AI from the outset would be a reasonable and sensible option. For this to be functional we must mandate this action in some capacity, just as privacy by design was introduced and mandated via GDPR. Seeing as embedding principles within a device or system from the earliest stages of development, the onus here would essentially fall within the hands of the manufacturer. As a result, in order for 'transparency by design' to work, it would make sense to begin by considering the development of industry standards in this space that would encourage uptake of this concept (the role of industrial standards will be considered in further on in this chapter).

5.3.1.2 Explainability

Explainability is sometimes regarded as its own separate ethical AI principle, however for the purpose of this thesis, it is considered in conjunction with transparency. This is because transparency and explainability overlap with one another; for an AI to be explainable, there must be some degree of transparency. For example, a simple way to increase explainability within AI might be to make clear which factors were taken into consideration by an algorithm used within an automated decision-making process. By making the system more transparent, it becomes more explainable. In this sense, it would be reasonable to suggest that a by-product of introducing 'transparency by design' would mean that incorporation of the principle of explainability could be achieved to some extent.

Most artificially intelligent systems are not able to comprehensively explain to users how a given autonomous decision was made.[551] This will proportionately affect the trust one puts into the system, as we cannot fully understand how a decision was reached. Lack of explainability and understanding of the functioning of most artificially intelligent systems is

[550] M. Boden, J. Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, P. Newman, V. Parry, G. Pegmanz, T. Rodden, T. Sorrell, M. Wallis, B. Whitby, A. Winfield, 'Principles of Robotics: regulating robotics in the real world' (2017) *Connection Science* 29(2) 124-129
[551] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G-Z. Yang, 'XAI – Explainable Artificial Intelligence' (2019) *Science Robotics* 4(37)

one of the primary causes for concern and forms the basis of most urgent calls for governance and regulation in this space.[552]

Therefore, explainable AI would give a user the ability to understand, manage and control the system, and better understand how certain decisions were reached.[553] By having more explainable AI, we can ensure that we are truly reaping all of the benefits that we can. If the functions performed by an artificially intelligent system are more explainable, then this will likely mean that we are able to identify new methods and strategies used by the system within critically important fields such as healthcare, and this could help us to make breakthroughs within research and development.[554] Any advancement in our ability to understand the functioning and actions of an artificially intelligent system will improve our relationship with this technology indefinitely.

Although, it is also salient to consider the concept of 'interpretability'. As per Gunning et al, explainability can be full or partial, and very much depends upon the nature of the AI system in question.[555] Interpretability is very much dependent upon the ability of an individual to understand the information being relayed to them by the system itself, which therefore presents us with another issue. Due to this, it very difficult to measure how comprehensible an artificially intelligent system is; the extent to which a system is sufficiently explainable depends on the level of technical understanding a person has, and their ability to interpret the information relayed to them.[556]

Despite these difficulties, explainability within AI remains a principle that is integral to improving overall trust in AI by ensuring verifiability can be achieved.[557] If we can ensure that the key principle of transparency is embedded within AI by design then we would by default improve the state of explainability also. However, the question remains as to if we can do more to incorporate the principle of explainability within AI regulation.

Education and skills

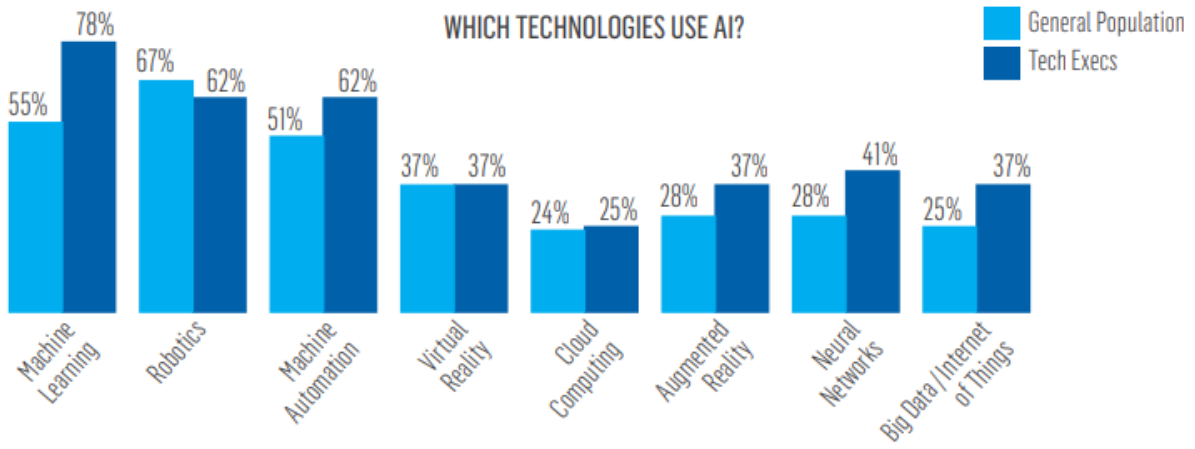
One potential solution that would allow us to improve explainability in AI, is to improve the current state of education and training surrounding AI. Here something similar to the approach taken by the UK government with regards to their National Cyber Security Skills

[552] J. Newman, 'Explainability won't save AI' (brookings.edu, 19 May 2021) <https://www.brookings.edu/techstream/explainability-wont-save-ai> accessed 02/08/2021
[553] Ibid
[554] W. Samek, K-R. Muller, 'Towards Explainable Artificial Intelligence' in W. Samek, G. Montavon, A. Vedaldi, L. Hansen, K-R Muller (eds) Explainable AI: Interpreting, Explaining and Visualising Deep Learning (Springer, 2019) 5-22
[555] Ibid n551
[556] Ibid n551
[557] Ibid n550

strategy is proposed.[558] In this context, the government recognised quite clearly the cybersecurity risks presented by the emergence of new technologies (including AI), and that individuals with sufficient understanding of these risks, and the skills necessary to prevent them, were scarce.

We have a growing community of talented and extremely knowledgeable individuals working within the realms of artificial intelligence. However, the general population's understanding of AI in general, how it is used and how it works could most definitely be improved.[559]

Figure 7: Technologies that use AI[560]



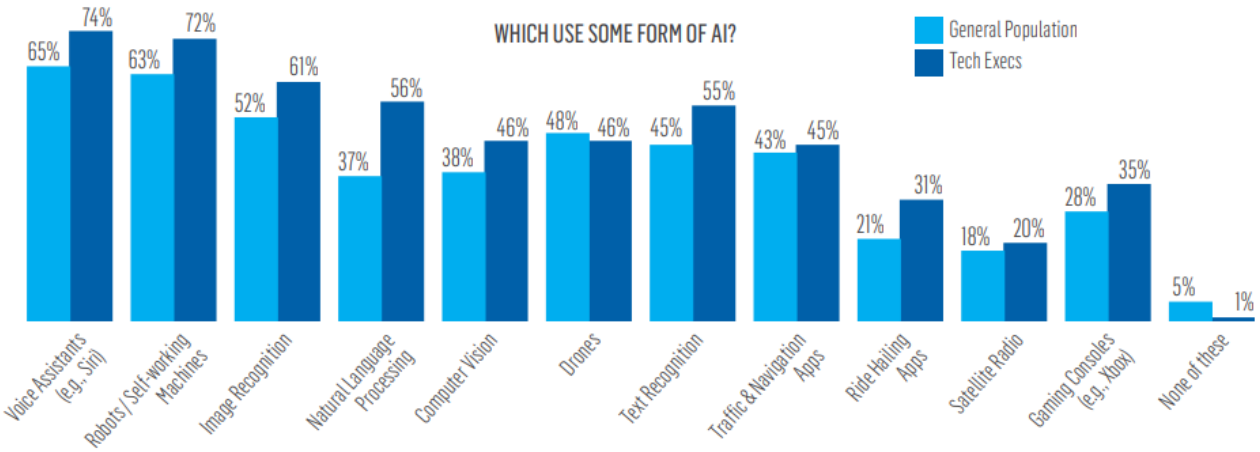
Source: Edelman, '2019 Edelman AI Survey' (2019) https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_AI_Survey_Whitepaper.pdf?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosfutureofwork&stream=future, accessed 05/08/2021

[558] Department for Digital, Culture, Media and Sport, 'Initial National Cyber Security Skills Strategy: increasing the UK's cyber security capability – a call for views' (*GOV.UK,* May 3 2019) <https://www.gov.uk/government/publications/cyber-security-skills-strategy/initial-national-cyber-security-skills-strategy-increasing-the-uks-cyber-security-capability-a-call-for-views-executive-summary> accessed 03/08/2021

[559] Edelman, '2019 Edelman AI Survey' (*Edelman.com,* March 2019) <https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_AI_Survey_Whitepaper.pdf?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosfutureofwork&stream=future> accessed 05/08/2021

[560] ibid

Figure 8: Technologies that use some form of AI[561]



Source: Edelman, '2019 Edelman AI Survey' (2019) https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_AI_Survey_Whitepaper.pdf?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosfutureofwork&stream=future, accessed 05/08/2021

Both figures 7 and 8 are taken from the 2019 Edelman AI Survey which was carried out in the summer of 2018 in the US. The above figures demonstrate the differences in understanding between tech executives and the general population regarding the various uses of AI. In summation, both the general population and tech executives understand that artificially intelligent technologies are used within voice assistant software and more obvious devices such as self-working robotic machines, however, both groups struggled to identify that AI is used within less obvious devices such as gaming consoles and taxi apps such as Uber.

Therefore, it is reasonable to say that both the general population and even those working within the tech industry understand that AI is being used within many widely available devices, yet there is a definite lack of understanding in how it works.[562] It is evident therefore that a broad understanding of how AI works is lacking within the general population and beyond, therefore by improving this understanding to a certain degree, we would likely have more success in deploying more explainable AI.

A quick online search for "improving AI skills" reveals a number of links to webpages and online courses that aim to help individuals gain 'AI skills' primarily to assist in securing a job within the tech industry.[563] However, there is little targeted at helping individuals to gain a

[561] Ibid
[562] Ibid
[563] For example, see the Saïd Business School (University of Oxford) 6 week, online Artificial Intelligence programme targeted at understanding AI for business <https://oxford-onlineprogrammes.getsmarter.com/presentations/lp/oxford-artificial-intelligence-programme>

high-level, general understanding of AI and its uses within our everyday lives. Accordingly, it would be reasonable to suggest that as a part of a durable and effective regulatory system, committing resources to help improve the high-level comprehension of AI and uses amongst the general population would be beneficial. A greater sense of understanding amongst the general public would help to enhance any efforts made by manufacturers to establish more explainable AI.

5.3.4 Concluding remarks on transparency

Transparency should act as cornerstone in any regulatory response to AI, and there are several ways in which these principles can be embedded within AI. The next part of this chapter considers more closely the actual methods by which we can build a regulatory model, with these principles at its centre.

Therefore, we will consider the various forms that legislation can take and the factors that must be considered when introducing a piece of legislation (particularly one that governs modern technologies), and we will also look more closely at the role of industry standards within regulations and legislation (again, specifically the merits of using them when tackling the issues presented by modern technologies).

5.4 Legislation

When considering how we can regulate in a particular area, the most obvious solution is to legislate. Legislation, regulation, and centralised governance allow public authorities to influence the way we live our day-to-day lives; they dictate the rules of public life and steer our behaviours.[564] Legislation is a powerful tool and can be conceived in a relatively short amount of time in order to make particular requirements mandatory, or certain behaviours and actions prohibited by law. For context, within in average parliamentary session (typically 12 months starting in Spring), the UK government can usually bring up to 30 bills before Parliament as part of a wider legislative programme, and these bills are assessed on their state of readiness and their priority.[565] Therefore, using formal legislation as a way to regulate is clearly both readily accessible to government departments, and also achievable

accessed 25/07/2021. And see also webpages such as the following which contains information on the 'Top 10 AI skills and how to get them' <https://www.techrepublic.com/article/here-are-the-10-most-in-demand-ai-skills-and-how-to-develop-them/> accessed 25/07/2021. This page focuses on 'skills' such as machine learning and data mining, as opposed to helping to enhance the high-level understanding of the general public with regard to AI.

[564] P. van Zwanenberg, A. Ely, A. Smith, *Regulating Technology: International Harmonization and Local Realities* (Routledge, 2013)

[565] Cabinet Office, 'Guide to making legislation' (*GOV.UK*, 14 July 2017) < https://www.gov.uk/government/publications/guide-to-making-legislation> accessed 05/05/2021

in a relatively short timescale (if a persuasive case can be made, and readiness demonstrated).

With specific regard to legislation intending to govern AI, there are a number of aspects that need to be considered such as: the flexibility and future-proofing ability of the legislative measures, the likelihood of the measures creating barriers to trade, and the capacity for alignment and international cooperation that the legislation will enable. Taking these variables into account is vital and should be considered during the rigorous 'pre-bill introduction' assessment.[566] As a part of this process, (which includes developing policy objectives, completing an impact assessment, undergoing independent scrutiny, and finally receiving policy clearance[567]) points like the ones listed here will need to be analysed in some depth.

Therefore, this section will consider the components necessary to create a fully functioning piece of legislation that could successfully govern AI; it will consider how we can go about establishing a future-proofed, non-obstructive and cooperative piece of legislation.

5.4.1 International alignment and trade considerations

By choosing to implement legislation, the opportunity for international harmonisation and alignment also arises; in some cases, this alignment is a requirement (e.g., within the EU, member states must align with the overall legislative agenda of the EU), but most of the time it is within the interest of countries to align their legislative proposals primarily for trade and monetary purposes. For example, most pieces of legislation must undergo World Trade Organisation (WTO) notification in order to assess the likelihood of the legislation creating barriers to trade.

Therefore, when considering the creation of robust legislation that will adequately manage the uses of AI, we must closely examine the impact that any proposed legislation might have upon the continuing operation of the industry; how might legislative measures affect development, production and the profitability of AI-based systems and products? For an idea of the scale of the industry in question; in 2019 it was reported that during the previous year over $5 billion was invested into just 1,400 sales and marketing companies that deal specifically in AI.[568] The investment seen here is only reflective of a small number of

[566] Department for Business, Energy and Industrial Strategy, 'The Better Regulation Framework' (2020) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/916918/better-regulation-guidance.pdf> accessed 20/04/2021

[567] Ibid

[568] P. Roetzer, 'Funding for AI Sales and Marketing Companies Exceeds $5.2 Billion' (*Marketing Artificial Intelligence Institute,* 8 Jan 2019) <https://www.marketingaiinstitute.com/blog/funding-for-ai-powered-sales-and-marketing-companies-exceeds-5.2-billion> accessed 30/04/2021

companies, in just one category of business that deals in the sale of AI. AI is becoming an increasingly lucrative business, with organisations such as ITU (the International Telecommunications Union) estimating that by 2030 AI could add around 16 percent to global GDP (which is around $13 trillion). As such, the impact that any legislative proposals might have upon this fruitful industry must be considered seriously: manufacturers and organisations do not want to implement and will not support regulations that may harm their ability to make profit and trade easily across borders (and the UK will want to remain a desirable location to do business).

Therefore, concepts of legislative alignment, cooperation, convergence and mutual recognition are incredibly important to consider when crafting legislation in any area, and even more so when attempting to regulate modern technology.[569] Fortunately, this now appears to be obvious to lawmakers across the world; e.g., following the European Commission sharing their proposed AI Act that aims to make clear to developers and users requirements and obligations necessary for specific AI applications.[570] The introduction of these 'harmonised rules' would provide for alignment across EU member state countries, but would also drive the desire for other nations to adopt similar regulations and rules, likely in an equally similar manner in order to prevent barriers to trade.

Similarly, no country in the modern world is truly self-sufficient; we live in a world of open economies and so operating across borders is entirely necessary, especially when it comes to modern technology.[571] As such, it is clear that any legislative measures in this context must utilise all options available to create minimal barriers to trade and be easily applicable across borders. However, regulations for modern technology often contain technical provisions, and there is often discomfort with this amongst bodies such as the EU and the WTO. Due to this discomfort, considerable delays can be placed upon legislative measures that contain technical provisions, to give these organisations the time to closely consider the impact that these technical provisions might have on international trade.

Interestingly, whilst working as a part of the Secure by Design team (with the Department for Digital, Culture, Media and Sport's Cybersecurity Directorate) on the Product Security and Telecommunications Infrastructure Bill, this exact issue was found. The Bill needed to

[569] D. Lawrence, 'Dynamic Alignment and Regulatory Cooperation between the UK and the EU after Brexit' (2019) Trade Justice Movement <https://www.tjm.org.uk/documents/briefings/TJM-Dynamic-Alignment-and-Regulatory-Cooperation-after-Brexit.pdf> accessed 20/04/2021
[570] European Commission, 'Regulatory framework proposal on Artificial Intelligence' (European Commission, 22 April 2021) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 24/04/2021
[571] G. V. Vijayasri, 'The Importance of International Trade in the World' (2013) International Journal of Marketing, Financial Services & Management Research 2(9) 111-119

contain certain technical provisions, however, including these provisions within primary legislation would mean potentially placing significant delay on the Bill which would postpone it from entering into force. We found that despite leaving the EU, EU notification was still an issue also due to our desire for the legislation to also be applicable in Northern Ireland (which would almost certainly be the case for any legislative measures for AI introduced in the UK as well).

As a result, it was decided to include the technical provisions within a batch of secondary legislation, so that the primary legislation (which contained the policy intentions of the Bill, and its general purpose and structure) could pass relatively smoothly without considerable delay. The merits to this were two-fold; firstly, it would allow us introduce the Bill as soon as possible, and to give manufacturers and other economic actors the ability to adjust their practices with plenty of time prior to enforcement of the legislation, and secondly it would allow the team to fine tune the technical provisions that would sit within the secondary legislation whilst still warning industry of what was to come.

Thus, it is advised that a similar approach be adopted within legislation introduced to govern AI. This approach would allow for legislation that has minimal impact upon international trade, forewarns manufacturers and economic actors as to the requirements that will be expected of them in due course, whilst giving policy and lawmakers the opportunity to fine-tune any technical provisions that will ultimately form a part of the legislation.

5.4.2 Future proofing

In a similar vein, when drafting a piece of legislation that has modern technology at its heart, the longevity of the legislation must be at the forefront of our minds, and as such we must ensure that it is future proof. According to the OECD (Organisation for Economic Co-operation and Development), this means "looking at the potential scale of effects in the long term… and emphasising the heterogeneity of choices".[572] The emphasis here on the "heterogeneity of choice" is rather interesting and promotes the idea that future proofed legislative efforts are likely most successful when combined with other equally effective regulatory methods. So, it is clear that any kind of legislative response to AI must be 'future-proof', any legislative measures introduced must be flexible and capable of adaptation.

[572] OECD, 'Regulatory Policy in the Slovak Republic: Towards Future-Proof Regulation' (2020), *OECD Reviews of Regulatory Reform* (OECD Publishing, Paris) <https://www.oecd-ilibrary.org/sites/94d061e5-en/index.html?itemId=/content/component/94d061e5-en> accessed 01/05/2021

As succinctly put by Jackson:

> "Creating a regulatory framework capable of accommodating all of the ethical dilemmas thrown up by this rapidly shifting terrain undoubtedly presents one of the most important and difficult tasks for law in the twenty-first century."[573]

Despite Jackson's work being primarily focused on the relationship between law, technology, and autonomy with regards to reproductive rights, her assertion still rings true. Creating a regulatory framework that can adapt to the 'rapidly shifting terrain' of technical development in this field, to keep the law from becoming obsolete, is both essential and incredibly difficult to achieve.

The definition of future proofing as used by Rehman and Ryan is particularly insightful, they state that "future proofing is an interdisciplinary perspective that offers a systemic framework to deal with future requirements and uncertainties while accommodating innovation".[574] Future proofing also remains an issue within other fields such as architecture, engineering, and medicine.[575]

So how might we go about ensuring that legislation intended to govern AI is future proof? The key here is making the future law capable of change in the least disruptive way possible. It is proposed that the easiest way to do this is by introducing legislative powers from the outset that will allow either the relevant individual, or enforcement body, to perform functions in the future that might not be necessary right now. A good example of this would be to include a power within the legislation at the outset that allows for the Secretary of State (in the UK) to introduce further technical requirements to the legislation as and when they become appropriate, and likewise the power to remove those that are no longer necessary. This allows for a good degree of flexibility that supports any future technical changes and allows the law to evolve with them.

5.4.3 Concluding remarks on legislation

The concepts discussed in this section, namely the international aspects of legislation and future proofing are of the utmost importance if we are going to make a sound and effective piece of legislation. Alongside the principles discussed earlier in this chapter, these legislative concepts will ensure regulatory success. Although, it is reasonable to suggest that legislation alone is not enough to tackle the issues that come with regulating AI, and as such

[573] E. Jackon, Regulating Reproduction: Law, Technology and Autonomy (Bloomsbury, 2001)
[574] O. U. Rehman, M. J. Ryan, 'On the Dynamics of Design of Future-Proof Systems' 25th Annual INCOSE International Symposium (2015) DOI: 10.1002/j.2334-5837.2015.00050.x.
[575] S. Ranchordas, M. Van-t Schip, 'Future-Proofing Legislation for the Digital Age' in S. Ranchordas, Y. Roznai (eds) Time, Law and Change (Hart, 2020)

a combined approach is likely to be more suitable, this would include utilising both legislation and industry standards.

5.5 Industry standards

Utilising industry standards within legislative agendas is common practise, particularly when the given legislation deals with a technical subject (such as AI). Many organisations choose to implement these standards to show that they are employing best practise within their industry, and that they are conforming with reputable, recognisable, and accepted norms. Within the UK there are several recognised standards development organisations (SDO's), and standards produced or adopted by these organisations have the capacity to become designated standards within UK law, these include:[576]

- British Standards Institute (BSI)
- European Committee for Standardisation (CEN)
- European Committee for Electrotechnical Standardisation (CENELEC)
- European Telecommunications Standards Institute (ETSI)
- International Organisation for Standardisation (ISO)
- International Electrotechnical Commission (IEC)
- International Telecommunication Union (ITU)

It is worth noting that standards developed by these SDO's are not just applicable in the UK, they are applicable all over the globe, by any government or organisation wishing to adopt the standard as a part of their legislation, or company best practice. This means that industry standards often provide internationally accepted baseline sets of standards for a given subject, yet unless they are incorporated or mandated by legislation (as a designated standard) they will not be legally binding.

By definition, a designated standard is "a standard, developed by consensus, which is recognised by government in part or in full by publishing its reference on GOV.UK in a formal notice of publication."[577] Designated standards are incredibly useful for a number of reasons; they are key policy tools and allow legislators lacking in technical expertise to effectively create adequate regulations, they prevent legislation from being too restrictive and promote

[576] Office for Product Safety and Standards, Department for Business, Energy & Industrial Strategy, 'Designated standards – guidance' (*GOV.UK,* 6 January 2021) <https://www.gov.uk/guidance/designated-standards> accessed 15/06/2021
[577] Ibid

flexibility, and they also allow manufacturers and other economic actors to play a direct role in deciding how they are to be governed.[578]

Therefore, within our proposed regulatory agenda, making use of technical industry standards that sufficiently cover the principles already discussed within this chapter would be well suited, and would effectively support any legislative measures put in place.

The issue remains however, there has been very little solid development in the AI standards space, and at present the organisation that appears to be leading on this work is ISO via their JTC 1 / SC 42 committee on artificial intelligence.[579] In particular, they appear to be working on a suite of AI standards that tackle different aspects of AI such as trustworthiness, systems engineering and data.[580] However, if we take trustworthiness as an example, at present the only available 'standard' is a technical report (which does not have the same influence or standing as an actual industry standard). The only actual standards that have been adopted and created by this committee so far are on 'big data' and measuring classification performance of machine learning models. To add to the issue, ISO charge for these standards as opposed to organisations such as ETSI that make their standards available for free. This is yet another hinderance as small and medium sized enterprises will likely struggle to implement standards that they have to pay for, which undermines their very intention.

Likewise, NIST (National Institute of Standards and Technology, a part of the US Government Department of Commerce) also have their sights set on developing standards in the AI space, yet again, there are few tangible outputs from this work programme as of yet.[581] It would appear as though, many standards development organisations are very aware of the risks and needs to create standards in this space, however little real progress has been made at present. With this being such a new and fast changing field with a variety of different aspects to consider ranging from cybersecurity risks to more ethical, moral hazards, creating adequate standards here will be a difficult task but a necessary one.

Despite this seemingly slow process, a clear benefit here is the ability for almost anyone with an interest or a role within the industry to become a member of a standards development

---

[578] L. Degallaix, M. Eliantonio, 'The use of standards in legislation and policies' (2018) *European Environmental Citizens Organisation for Standardisation* <https://ecostandard.org/wp-content/uploads/2018-06-11-The-use-of-standards-in-legislation-and-policies-ECOS-discussion-paper.pdf> accessed 12/06/2021

[579] ISO, 'Towards a trustworthy AI' (*ISO.org,* 7 July 2020) <https://www.iso.org/news/ref2530.html> accessed 10/06/2021

[580] Ibid

[581] NIST, 'Technical AI standards' (*nist.gov,* 6 August 2021) <https://www.nist.gov/artificial-intelligence/technical-ai-standards> accessed 10/08/2021

organisation and help to get the ball rolling. Within ETSI for example, there are a great variety of members including government bodies, manufacturers, both public and private research bodies, consultancies and universities to name a few.[582] These 900 members also come from countries all around the world, which reflects the strengths of standards organisations such as ETSI, they have considerable global reach and the capacity to encourage international cooperation on contemporary issues such as the ones presented by AI.

Therefore, despite having a seemingly slow start without much solid, tangible progress in the space so far, the role that standards can and will play in the regulation of AI is undeniable and should not be underestimated.

5.5.1 Making use of standards in this space

Now that we have discussed the merits of making use of standards to support legislation, it is worthwhile considering what exactly we would like to see included within a standard that would make meaningful impact in this space. Earlier in the key principle of transparency was discussed, and how it is vital that any regulatory response to AI has this principle incorporated and embedded within it. In particular, in order to successfully incorporate the principle of transparency within our regulatory response, standards could most certainly be utilised here.

It was proposed earlier in this chapter that to ensure that transparency is included within an artificially intelligent system, we could introduce 'transparency by design'. This would mean ensuring that the system is built in such a way that it is sufficiently transparent from the very beginning. Inspiration was drawn from the approach taken by the UK government's Secure by Design work that aims to embed security principles within consumer IoT devices from the very outset. Like the Secure by Design work currently being translated into UK-wide legislation, industry standards played a central role in ensuring that the relevant technical requirements and measures were adequately reflected within the regulations.

So, what would the ideal standard look like? We would need a standard that highlights the importance of transparency as a key underpinning principle of AI, and its impact upon trustworthiness. And within this standard, we would like to see provisions that detail the ways in which we can promote transparency during the life cycle of the system, and ways in which we can mitigate any potential vulnerabilities and their impacts.

[582] ETSI, 'Membership of ETSI' (*etsi.org,* 2021) <https://www.etsi.org/membership/members> accessed 10/08/2021

Again, taking the Secure by Design work and its use of EN 303 645 as an example, this standard lists thirteen provisions that intend to improve the state of security of consumer IoT devices, with the onus primarily put upon the manufacturer. These provisions include banning default passwords within devices, implementing a means to manage vulnerability disclosure, and keeping software updated.[583] The other provisions are relatively similar in nature, and are a mix of procedural and technical requirements which adds strength to the standard. At present, the UK legislation is only mandating the first three requirements from EN 303 645, as listed here, with scope to introduce the remaining ten provisions over the coming years.

Therefore, and in a similar vein, if we were to create a standard that aims to introduce the principle of transparency to AI from inception, it would also be worthwhile having a combination of provisions that are both procedural and technical in nature. To begin with, introducing a provision that requires increased algorithmic transparency would be beneficial; this is not to say that we must entirely remove the whole concept of the 'black box' at once, but more so to begin moving down that path. By having a provision that requires some level of disclosure regarding the factors and variables taken into consideration by the AI system when reaching a decision would most definitely be valuable.

It would also be valuable within this standard to have a provision that allows for instances of suspected discrimination or bias within an AI system to be easily disclosed. Similar to provision 5.2 of EN 303 645, requiring manufacturers and other organisations to have an adequate procedure in place for researchers and others to report instances in which they have found that an AI system has acted in a biased or discriminant manner, would promote more awareness of this issue. By including a provision that deals with this issue, manufacturers and those developing AI will be encouraged to give more thought to the potential biases within their systems. Not only this, but in an effort to increase trust, the general public will be reassured that bias and discrimination is a factor that can be actively monitored by those in control of the systems that we are subjected to, and that transparency is a principle being pursued.

Another potential provision for inclusion that would ensure transparency is embedded within these systems by design is to ensure that developers are adequately controlling the data sets that they are using to train artificially intelligent systems. This would require developers to be aware of the impact that the data that they are using might have upon the potential

[583] ETSI EN 303 645 https://www.etsi.org/deliver/etsi_en/303600_303699/303645/02.01.01_60/en_303645v020101p.pdf accessed 22/11/2022

outputs. It would mean that manufacturers and developers are required to use diverse data sets that are truly representative of the society in which we live and ensure that they are not adding fuel to any existing inequalities.

These are three high-level requirements that could easily form the basis of an industry standard with the intention of promoting and ensure trustworthy AI, specifically via advocating for transparency. Likewise, by introducing the principle of transparency in this way, we can also begin to make AI more explainable too (although, as per the discussion earlier, standards would likely not be the solution to this exclusively).

5.5.2 Concluding remarks on the use of industry standards

It is clear that industry standards have a role to play in the regulation of AI, despite the current lack of robust standardisation work in this space. Although, we have few suitable standards at present, it is promising to know that standards development organisations are aware of the work that needs to take place here. Therefore, there is definite capacity for interested individuals from academia, industry, and government bodies to get involved with the development of a standard to fill the existing gap. These organisations encourage as much involvement as possible from a variety of backgrounds, and from individuals with varying experience, and as such outputs of this work will certainly be incredibly beneficial in the battle to regulate AI.

5.6 Conclusions

In this chapter, varying points and concepts that are key to creating the ideal regulatory response to AI have been examined. By assessing these methods, concepts, and models and their merits this chapter has highlighted several problem areas, and opportunities for development. This chapter therefore addresses the third research question set out within this thesis; what key regulatory principles are valuable and should be included in an ideal AI governance framework?

The next chapter in this thesis will consider more closely the shape that AI regulation might take by proposing a comprehensive method for AI governance and considering a number of essential governance criteria such as the existence of an adequate regulator, the target audience, and appropriate penalties for non-compliance with the regulation.

Chapter Six

A Proposal for AI Governance

6.1 Introduction

Drawing upon themes discussed within previous chapters of this thesis, this chapter will include suggestions for an AI regulatory framework, addressing the fourth and final research question set out in this thesis; what realistic and workable recommendations can be made to improve and secure the current state of AI regulation? This proposal will take inspiration from AI regulatory initiatives adopted and proposed by various states and organisations, legislative agendas in similar technological fields such as IoT, existing legislation and governance measures in field such as human rights, equality protection and product safety, and industry standards. It is a combination of these elements that the author suggests will amount to an adequate AI governance framework.

When piecing together a functioning regulatory framework, one must consider several key elements, who the regulator will be, what the target is, what the framework commands, and what the consequences of the regulations will be.[584] Therefore, this chapter will begin by featuring an overview of a proposed regulatory framework, and then will be further broken down into four primary sections as per the above regulatory elements.

6.2 Overview of the Framework

As established already in this thesis, AI regulation is necessary. And whilst work is being done at a global level to form adequate regulatory regimes and governance measures, many of the 'leading' approaches have several shortcomings or are relatively underdeveloped. This proposal for an AI regulatory framework aims to offer an effective solution to these shortcomings and present a number of coordinated governance measures that could provide a practical system for regulating AI.

At its heart, the framework features a risk-based approach for regulating AI (similar to the one proposed within the EU AI Act and the UN recommendations, albeit with amendments) and draws upon a number of regulatory measures discussed within Chapter Five of this thesis, including primary legislation, secondary legislation where necessary, and industry standards.

The aim of this proposed regulatory framework is multifaceted; overall, this proposal promotes the development, deployment, and use of 'safe' AI, and applies to economic actors involved in the AI life cycle, and ultimately users of AI (a group neglected by the EU AI Act

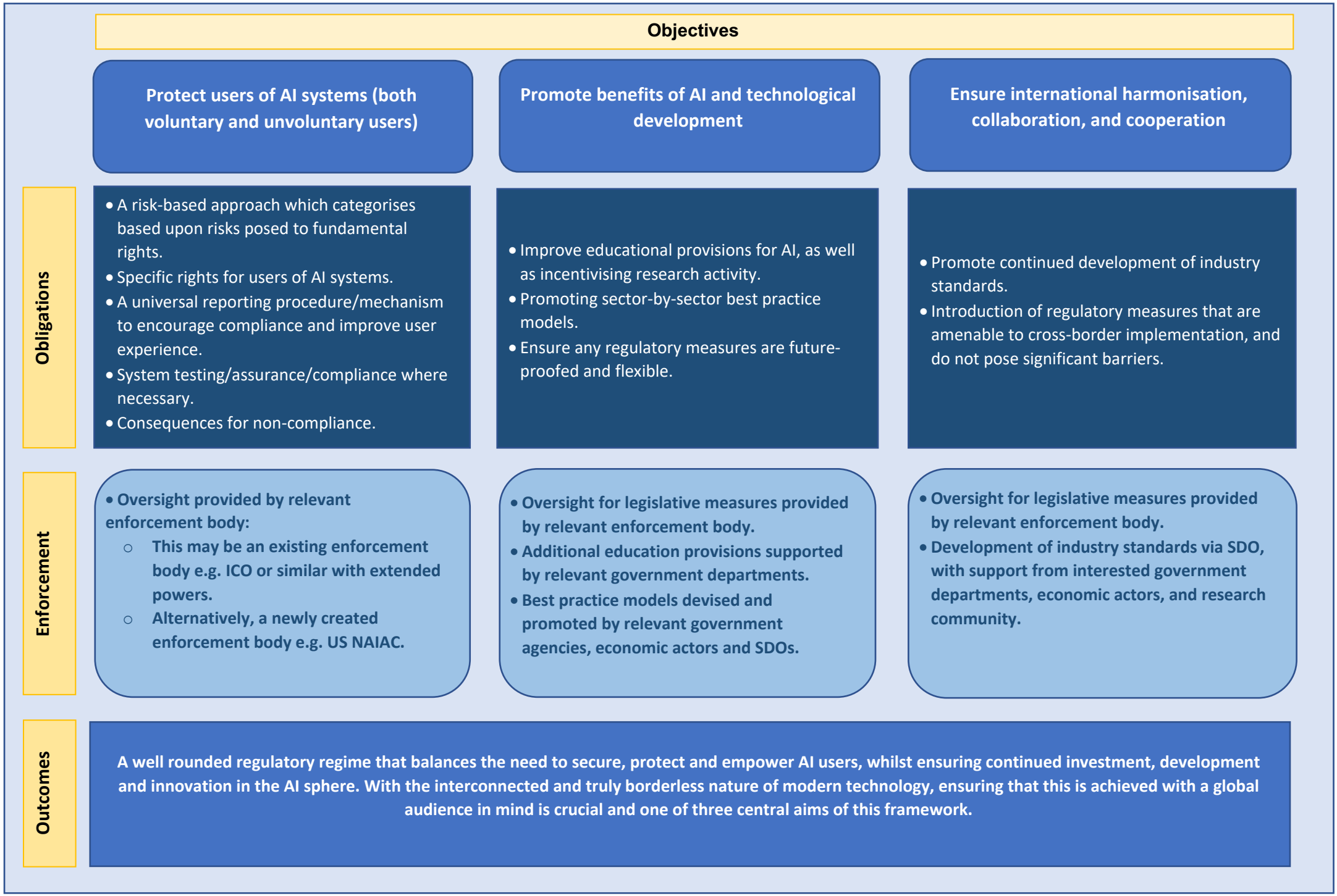
[584] C. Coglianese, 'Regulation's Four Core Components' (2012) *The Regulatory Review* < https://www.theregreview.org/2012/09/17/regulations-four-core-components/> accessed 20/06/2022

proposal). ‘Safe’ in this context can be understood as meaning AI that is to a reasonable extent explainable and transparent, accountability for harm is ensured, and the AI is fair (more detail on this will follow). In other regulatory proposals, the focus has often been on promoting trustworthy AI, yet this is often a concept that has proved rather difficult to adequately define.[585] Trustworthiness is somewhat of an umbrella term for several AI related concepts, such as bias or machine learning ability, some of which are more applicable to some applications than others.[586]

This therefore means that any regulatory system that centres around the concept of promoting trustworthiness in AI runs the risk of being rather uncertain and nonuniform in its application. As a result, it is suggested that a more reasonable term to use within the framework is ‘safe’; this still encapsulates the concepts that trustworthiness alludes to, yet it is more consistent in its application e.g., it is reasonable to say that all AI applications should be safe.

[585] J. Harris, B. Ammanath, ‘Defining trustworthy AI’ (2022) *Towards Data Science Podcast Transcript* <https://towardsdatascience.com/defining-trustworthy-ai-234a97c39035> accessed 20/06/2022
[586] Ibid

Figure 9: Regulatory Framework Overall Aims



The previous figure is intended to give a brief overview of the proposed regulatory framework, including its overall policy objectives, obligations necessary to achieve those objectives, how these obligations will be enforced and what the overall outcomes of these policy objectives, obligations and enforcement measures will be. The following sections of this chapter consider in some detail how each of these regulatory elements will function in practice and provide reasoning for their inclusion in this proposal.

6.3 The Regulator

One of the most important factors for any impending regulation is to have an adequate regulator in place; after all, rules are useless if there is no singular body available to enforce them. This therefore poses the question, what makes a good regulator? There are a number of desirable characteristics that a good regulator should have, and within this section the most preferable characteristics will be considered, as well as the capability of existing bodies, agencies, and administrations to step up as a central regulatory body for AI.

This section uses primarily UK regulators as examples to demonstrate points relating to the role of an AI regulator. However, with regards to who or what the ideal AI regulator might look like, this will likely depend on the specific jurisdiction in question, e.g., in reference to domestic AI regulation a national regulator might be most suitable, whereas for regional or international regulatory measures an international regulator may be most suitable.

6.3.1 Crucial regulatory principles

In 2006, the World Bank published a handbook for evaluating regulatory systems, and within this handbook they included a useful list of principles necessary for an effective economic regulator.[587] The author has selected the three most important of these principles, which will be discussed in this section with reference to how they should apply to an AI regulator.

First and foremost, in this list is ‘independence’. This is a valuable principle and one which is applicable to any potential AI regulator; it goes without saying that the regulator should be able to make reasonable decisions without having to seek confirmation from another body or agency first.[588] This is the same argument put forth for the independence of the judiciary for example; it is necessary that the judiciary operates independently from the executive and legislative branches of government in order to maintain integrity, impartiality and remain free of improper influence in order to uphold the countries democratic values. In applying this to a regulator of AI, it is also vital that the regulator is able to remain independent from influence

[587] A. C. Brown, J. Stern, B. Tenenbaum, D. Gencer, *Handbook for Evaluating Infrastructure Regulatory Systems,* The World Bank (2006) p. 59-63
[588] Ibid

of government agendas, and equally remain impartial to the desires of the tech community, both of which will likely present some competing values.

Independence is also flagged by the UK government as foundational when establishing a functional assurance ecosystem.[589] The independence of the assurance provider in this ecosystem would ensure that trust in the systems undergoing assurance is justified.[590] Whilst independence is essential for these bodies, it is worth acknowledging the close relationship that they typically have with one or more government departments; this is an unavoidable factor. The Information Commissioners Office (ICO) for example is the independent regulator for data protection and freedom of information in the UK.[591] Despite their independence in governing data use in the UK, they are sponsored by the Department for Digital, Culture, Media and Sport (DCMS) meaning that the body has an inevitable close relationship with the government department due to regulating many of its legislative and regulatory initiatives; however, the ICO still remains independent.

The second principle essential for an AI regulator is transparency, specifically with regards to the public.[592] As with AI itself, it is vital that rules, procedures, and decisions made by the regulator should be transparent, meaning that the public are able to understand these processes and easily access relevant documents. This will go some way to ensure that the regulations themselves are truly human-centric, a short-coming of the EU AI Act proposal (discussed within Chapter Four), and therefore put the actual 'user' at the heart of the regulations. As per relevant data protection regulations (e.g., GDPR), ensuring that the user (or data subject in the case of GDPR) has rights to access information regarding how their information is used, and the ability to report wrongdoing is essential to the functioning of the regulations and the success of the regulator.[593]

The Financial Conduct Authority (FCA), the regulator of financial services firms and financial markets in the UK, uphold transparency as one of their key regulatory principles as it allows them to be scrutinised by interested parties, e.g. government, the firms they regulate and the general public.[594] The same principle should be applied to an AI regulator; users of AI systems should be at the heart of the regulations, and therefore providing these individuals

[589] Centre for Data, Ethics and Innovation, 'The roadmap to an effective AI assurance ecosystem' (GOV.uk, 2021) < https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem> accessed 20/06/2022
[590] Ibid
[591] ICO, 'Who we are' (ico.org.uk, 2022) < https://ico.org.uk/about-the-ico/who-we-are/> accessed 20/06/2022
[592] Ibid n4
[593] Ibid n25
[594] Financial Conduct Authority, 'Transparency' (fca.org.uk, 2022) < https://www.fca.org.uk/about/transparency> accessed 20/06/2022

with the ability to easily view and challenge the decisions and processes of the regulator is necessary.

The final regulatory principle of particular importance is the clarity of roles.[595] Ensuring that the role the regulator is to play in the regulation and governance of AI is clear is pivotal. By making clear which specific role the regulator is to fulfil in law should avoid confusion as to which body bares the responsibility for regulating particular factors, duplication of functions and prevents incorrect messages being communicated to stakeholders and the public.[596] This is a relatively simple to do; for example, Ofcom are the independent regulator of broadband, phone services, tv and radio in the UK, they are therefore listed as the regulator for these services in a number of Bills, with their powers clearly defined in each.[597] Ofcom's regulatory duties include ensuring communication services are available to the public, ensuring that there is diversity in tv and radio, that these broadcasts are free of offensive content, and postal services are available to all UK addresses.[598] However, it is stated clearly within both legislation and guidance provided by the regulator that they are not responsible for regulating individual disputes between a provider and members of the public, the BBC World Service, post offices or what people write and post on the internet.[599]

With regards to AI regulation, it is essential that any governing body must be given clear indication as to what falls within their remit and what does not. For example, it is relevant for the regulator to be charged with governing instances in which there is a report of wides-spread non-compliance, however it may be less relevant for them to govern the implementation of an AI-enhanced curriculum (this duty is more likely to sit with the Department for Education or equivalent).

6.3.2 Current state of AI regulatory bodies

With the most pertinent regulatory principles in mind, it is relevant to consider the current state of AI regulatory bodies on a global scale, i.e., are we creating new regulatory bodies specifically to regulate AI or are we simply enhancing the powers and scope of existing bodies such as the ICO and Ofcom. There are arguments for against each of the above approaches, some of which will be explored within this section.

[595] Ibid n587
[596] Ibid n587
[597] Department for Digital, Culture, Media and Sport, 'World-first online safety laws introduced in Parliament' (GOV.UK, 2022) < https://www.gov.uk/government/news/world-first-online-safety-laws-introduced-in-parliament> accessed 20/06/2022
[598] Ofcom, 'What is Ofcom?' (ofcom.org.uk, 2022) < https://www.ofcom.org.uk/about-ofcom/what-is-ofcom> accessed 20/06/2022
[599] Ibid

From a UK perspective, it appears as though the ICO is positioned as an obvious regulator for AI. A number of the ICO's central priorities e.g., data protection and information rights have a huge overlap with many of the principles in need of regulation for AI. The ICO have also produced a number of AI guidelines including information regarding explaining the use of AI to individuals, AI and its relationship with data protection, and a data analytics toolkit.[600] Interestingly however, in their guidance on big data, AI, machine learning and data protection, the ICO stated that they believe existing legislation is capable of governing AI.[601] This view was supported by the House of Lords Liaison Committee in their 2020 report, in that GDPR appears to adequately deal with data protection concerns regarding AI.[602]

This is in contention with views already raised in this thesis (see Chapter Two for further detail); if this is the approach of the ICO, then their suitability as an AI regulator should be questioned at the very least. Whilst generally hailed as playing a successful role in the regulation of data protection in the UK, the ICO has been criticised for "…overseeing a regime that is not meeting its objectives either in fundamental rights or economic terms."[603] One example of such failing is the implementation of the Age Appropriate Design Code (AADC) by the ICO, which in fact had no real basis in law, and was deemed to impose significant burden upon e-commerce providers but was not subject to an impact assessment.[604]

Therefore, whilst the ICO is the most logical choice for an AI regulator in the UK at present, it is still worthwhile considering other options, and whether or not creation of a new regulator would be more suitable. The UK government recently commented on the future of AI regulation in the UK, and specifically who would be tasked with regulating AI.[605] It appears as though a sector-by-sector approach is being favoured, meaning that existing regulators will be relied upon to regulate AI use in their respective sectors. This will certainly save money and reinforces the UK's intention to regulate AI on a sector-by-sector basis.[606]

---

[600] R. Free, C. Kerrigan, B. Zapisetskayac, 'AI, Machine Learning & Big Data Laws and Regulations 2022, United Kingdom' (2022) *Global Legal Insights* < https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/united-kingdom#chaptercontent3> accessed 20/06/2022
[601] ICO, 'Big data, artificial intelligence, machine learning and data protection' (ico.org.uk, 2017) < https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf> accessed 20/06/2022
[602] House of Lords Liaison Committee, 'AI in the UK: No Room for Complacency' HL 196 2019-21
[603] V. Hewson, J. Turnbridge, 'Who regulates the regulators?' (2020) *Institute of Economic Affairs* < https://iea.org.uk/wp-content/uploads/2020/07/Who-regulates-the-regulators_.pdf> accessed 20/06/2022
[604] Ibid
[605] Ibid n244
[606] Ibid n244

With regards to the EU and its regulation of AI, the approach seems rather similar to the UK in that the intention is to regulate at Member State level by building upon 'already existing structures'.[607] This could therefore be interpreted to mean that Member States will initially be relying on already existing regulatory bodies, likely with additional powers granted and scope increased to include AI. Due to the structure of the EU, this would be supported at Union level via a newly created European Artificial Intelligence Board that would ultimately oversee the regulations.

Yet, as with the proposed Act itself, the information and guidance we are provided with regarding the proposed enforcement structure is vague; we are told little other than existing structures will be relied upon at Member State level, and that the European Artificial Intelligence Board will be chaired by the European Commission and will likely feature representatives from each of the Member States governing bodies.[608]

As discussed in some depth in Chapter Three of this thesis, the US' fragmented approach to AI regulation means that they have a number of committees and agencies specifically looking at AI related issues, but at present none specifically tasked with the regulation of AI in general (e.g. the National Security Commission on Artificial Intelligence).[609] However, due to the nature of the US' constitutional structure, it is unlikely that we will see one such body created, it is more likely that we will see a sector by sector approach to regulation similarly relying on existing structures for enforcement.[610]

As a result, we can see a trend appearing amongst the leaders in AI regulation; relying on existing governance structures and bodies and enhancing those with additional powers and scope is the most obvious choice for enforcement of AI regulations, as and when they arise. Although this is a trend that we are seeing with regards to enforcement, it does not necessarily mean that this is the most suitable option.

Taking the ICO as a case study, it was created in 1984 with the enactment of the Data Protection Act 1984, the first legislation of its kind.[611] At that point, there was need for a new regulator to deal with legislation in a relatively new field, to promote understanding of the new rules and encourage sectoral best practice.[612] As time went on legislation progressed

---

[607] Ibid n8
[608] M. MacCarthy, K. Propp, 'Machines learn that Brussels writes the rules: The EU's new AI regulation' (2021) *Lawfare* < https://www.lawfareblog.com/machines-learn-brussels-writes-rules-eus-new-ai-regulation> accessed 20/06/2022
[609] National Security Commission on Artificial Intelligence, 'About us' (nscai.gov, 2022) < https://www.nscai.gov/about/> accessed 20/06/2022
[610] See Chapter 4 for further detail
[611] ICO, 'Our history' (ico.org.uk, 2022) < https://ico.org.uk/about-the-ico/our-information/history-of-the-ico/our-history/> accessed 20/06/2022
[612] Ibid

and the ICO grew into the significant enforcement body that it is today, and the same argument could be put forth for AI.

As opposed to trying to fit the regulation of AI into an already existing mould (i.e., by relying on the ICO to regulate it in the UK), creating a new regulatory body specifically tasked with governing AI development, deployment, and use, could be much more beneficial in the long run. A similar approach was recently taken in the UK construction industry in which the government announced a new regulator would be established to carry out complaint investigation, market surveillance and other duties for construction products in the UK.[613] It is therefore possible, if deemed necessary, for a new independent regulatory body to be established to oversee new AI rules and regulations.

In the case of AI, it is arguable that the establishment of a new regulatory body would be more appropriate than reliance on an existing one such as the ICO in the UK. This is for several reasons; firstly, and as discussed already in this thesis, existing regulations such as GDPR and the Equality Act 2010 are not fully capable of regulating AI as they stand, yet those at the ICO are of the opinion that they are adequate.

Secondly, despite AI being compared to other 'related' tech trends such as IoT, virtual reality, blockchain etc,[614] it does have its differences, and these should be considered. Just because some of the issues we face due to AI overlap with other areas of tech, for example data protection and information sharing, does not mean that those issues should be approached in the same way. AI is unique as its abilities differ from other types of tech; most AI, and especially the more sophisticated forms of AI, has the ability to 'think' for itself, therefore changing its outputs and actions based on these observations.

As discussed already, this sets AI apart from its counterparts in tech and it is this uniqueness that poses such novel risks and issues that we are attempting to address via regulation. Therefore, for such an advanced and unfamiliar type of technology that is increasingly invading every part of everyday life, it is reasonable to suggest that we need a new regulatory body with specific expertise in this area to fully enforce and develop necessary

[613] S. J. Dobson, K. Ciclitira, 'A new regulatory regime and a new regulator: a new era for the regulation of construction products in the UK?' (2022) *Kennedys Product Safety Blog: In Safe Hands* < https://kennedyslaw.com/thought-leadership/blogs/product-safety-blog-in-safe-hands/a-new-regulatory-regime-and-a-new-regulator-a-new-era-for-the-regulation-of-construction-products-in-the-uk/> accessed 20/06/2022

[614] PWC, 'Eight emerging technologies and six convergence themes you need to know about' (2022) < https://www.pwc.com/us/en/tech-effect/emerging-tech/essential-eight-technologies.html#:~:text=They%20include%3A%20artificial%20intelligence%20(AI,pandemic%20accelerating%20emerging%20tech%20adoption.> accessed 20/06/2022

regulation and guidance; one that can dedicate its whole attention on the development and deployment of 'safe' AI without having to compromise.

It is however worth acknowledging the cost that would likely be incurred by creating a new AI regulator. For example, in 2016 the UK government announced without much notice that the regulatory body for governance of social workers would be changing to a new body at the estimated cost of around £15 million.[615] In reality, a previous change to the regulator in 2012 cost around £17.6 million; this cost wasn't for the creation of a new regulator, but rather the adaptation of an already existing one.[616]

In similar fashion, the new regulator for construction products was announced by the government in 2021 at an initial estimated cost of £10 million.[617] It is therefore clear that the creation of an entirely new regulator would be costly, and at present whether or not this would be a cost deemed appropriate is debatable but is worth considering, nonetheless.

6.3.3 Conclusions

As demonstrated in this section, there are several crucial regulatory principles that a regulating authority must embody. There are also two ways in which we could approach the regulation of AI; relying on existing regulatory bodies to do the job by granting them additional powers and scope or creating a new regulatory body altogether. It would appear as though the first option is the one being adopted by most nations interested in AI regulation, although this thesis contests there is more merit in choosing the second option and creating a new regulatory body.

It is suggested that a new enforcement body would oversee the general regulatory framework as suggested in this Chapter. It is without doubt that there would be some overlap with existing legislation enforced by other regulatory bodies, e.g., data protection issues governed by the ICO, product safety issues regulated by the Office for Product Safety and Standards (OPSS) in the UK. However, if the newly proposed regulator embodies the core principle of clarity of roles, then the potential conflict and overlap that may arise can be circumnavigated (primarily through precise wording of legislation).

---

[615] A. McNicholl, 'Social work's new regulator will cost millions. Who will foot the bill?' (communitycare.co.uk, 2016) <https://www.communitycare.co.uk/2016/01/27/social-works-new-regulator-will-cost-millions-will-foot-bill/>

[616] Ibid

[617] S. J. Dobson, K. Ciclitira, 'A new regulatory regime and a new regulator: a new era for the regulation of construction products in the UK' (kennedyslaw.cm, 2022) <https://kennedyslaw.com/thought-leadership/blogs/product-safety-blog-in-safe-hands/a-new-regulatory-regime-and-a-new-regulator-a-new-era-for-the-regulation-of-construction-products-in-the-uk/> accessed 20/06/2022

Again, whilst remaining independent, the regulator would collaborate closely with government departments (most likely DCMS and BEIS and the Department for Education (DfE) in the UK), whilst continuing to harbour relationships with the world's most prominent SDOs in order to promote international harmonisation of regulations. For the regulatory framework proposed here to be effective, we must also consider the target of such regulation with whom the regulator must work with closely.

6.4 The Target

Another essential component within any regulatory regime is ensuring the target of the measures is clearly identifiable.[618] It is imperative that we are able to easily determine who in particular the regulations apply to, and therefore which parties will suffer the consequences of non-compliance; in turn, this signifies clearly to the relevant parties what their obligations are and minimises the risk of non-compliance. This section considers the ideal target audience for the proposed regulatory framework. In doing this, the section is broken into two main sections; the first part considers the regulatory impact upon users, and the second part looks at the actions expected from economic actors.

6.4.1 Defining the 'user'

As discussed in Chapter Three of this thesis, one of the most obvious shortcomings of the proposed EU AI Act is that despite claiming that the regulations are human-centric and protective of fundamental rights enshrined in the likes of the European Convention on Human Rights or the Universal Declaration of Human Rights, they are predominantly targeted at manufacturers and other economic actors within the life cycle, and therefore neglect to cater to the end-user of the system.[619] Further to this, the proposed Act refers to the 'user' as a deployer of AI as opposed to using the general understanding of the term (e.g. the individual using the system itself). Any confusing use of terminology such as this will likely cause legal uncertainty and therefore impede upon the effectiveness of the proposed legislative measures.

The end-user of an AI system is ultimately going to be the individual that bears the impact and consequences of the outcomes of the system itself. This individual may have used the system out of choice (e.g., by buying an autonomous vehicle) or the system may have formed part of an unavoidable public service (e.g., visa application, or medical diagnosis). Therefore, the end-user is an undeniably important component within any regulatory regime;

[618] Ibid n584
[619] Ibid n8

as per this proposal, the end-user should be empowered by legislative measures and given the opportunity to play an active part in their AI-based experiences.

Therefore, it is pertinent to define what is meant by the term 'user' for the purposes of this thesis. Here, user is interpreted to mean the same as 'end-user' of an AI system, e.g., "the person or organisation that *uses* a product or service".[620] The user in this sense is therefore the individual that will typically be subject to the outcomes of the AI system, as opposed to an individual or organisation specifically involved in the development and/or deployment of the system itself.

This does not mean that manufacturers and other economic actors invested in the AI life cycle would be neglected by the regulations, but rather that those directly affected by AI outcomes (some of which may well be unintended) would be protected and empowered to seek recourse for an AI-based wrongdoing. One way in which this objective may be achieved is by mandating a reporting mechanism for AI systems within the regulations; this would therefore give AI users a voice to raise concerns and report instances of misconduct directly to those deploying and developing these systems.[621]

The approach suggested here is similar to the one included within the forthcoming Product Security and Telecommunications Infrastructure Bill (PSTI Bill) on consumer IoT (Internet of Things) devices.[622] This requirement provides a clear route for individuals to report vulnerabilities to the manufacturer of the product in question, allowing for security vulnerabilities to be resolved.[623] At present this type of reporting procedure is rather uncommon for consumer IoT devices, but it is believed that the mandating of this provision within law will vastly improve cybersecurity within the sector. Therefore, it is reasonable to suggest that a similar approach would work well for AI; by allowing for simple and transparent reporting of faults within AI systems would allow for any resulting risks to be minimised and for systems to be amended at pace.

---

[620] Cambridge Dictionary, 'End user' (dictionary.cambridge.org, 2022) <https://dictionary.cambridge.org/dictionary/english/end-user> accessed 25/06/2022

[621] Chapter Three

[622] Product Security and Telecommunications Infrastructure Bill, HL Bill 16 58/3

[623] Department for Digital, Culture, Media and Sport, 'Government response to the call for views on consumer connected product cyber security legislation' (GOV.uk, 2021) <https://www.gov.uk/government/publications/regulating-consumer-smart-product-cyber-security-government-response/government-response-to-the-call-for-views-on-consumer-connected-product-cyber-security-legislation> accessed 28/06/2022

Figure 10: Reporting Mechanism for AI Concerns and Outcomes

To aid in the implementation of a reporting mechanism as per the forthcoming PSTI Bill, the National Cyber Security Centre (NCSC) have produced a useful vulnerability disclosure toolkit for organisations to make use of when setting up their mandated disclosure processes.[624] This toolkit provides guidance regarding how an organisation should communicate the availability of their new reporting procedure, what information the reporter should be providing and what the organisation will do in response to receipt of this information, and how the reporter should report the information they wish to disclose (e.g. via a manned email address listed clearly on a dedicated page on the organisations website, or via a webform).[625]

Guidelines such as these provided by the NCSC further lessens the potential burden on manufacturers and economic actors when introducing new, rather simplistic regulatory measures. This is beneficial for all relevant parties; by relying on guidance produced and published by government departments and other related agencies to support the implementation of clear-cut regulatory measures such as the one proposed here means that more resources can be focused on implementing the more complex measures that will be discussed further on in this chapter.

Similarly, this reporting procedure could serve another equally important function, it could be used to report user feedback on the explainability of the system in question. As discussed already in this thesis, and throughout the literature in this space, the explainability of AI (which is intrinsically linked to transparency) remains to be one of the largest issues we face as AI development continues to develop at pace.[626] There is no clear solution to improving AI explainability for the general user, however, by using a reporting feature such as this would give users the ability to effectively communicate their experiences with a given system, and provide manufacturers with useful feedback that would allow them to develop increasingly transparent and understandable AI.

---

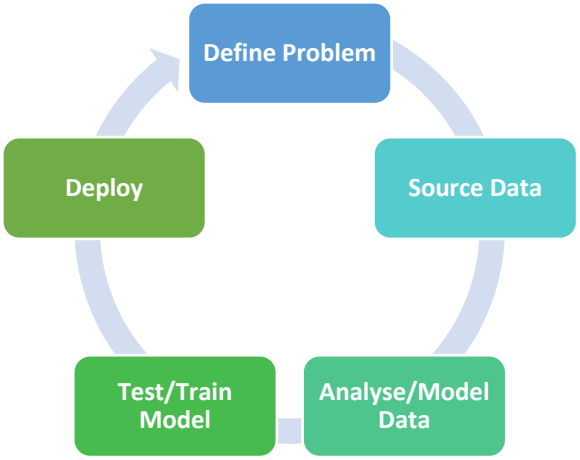
[624] National Cyber Security Centre, 'Vulnerability Disclosure Toolkit' (ncsc.gov.uk, 2020) <https://www.ncsc.gov.uk/information/vulnerability-disclosure-toolkit> accessed 28/06/2022
[625] Ibid
[626] S. Laato, M. Tiainen, A. K. M. Najmul Islam, M. Mantymaki, 'How to explain AI systems to end users: a systemic literature review and research agenda' (2022) *Internet Research* 32(7)

6.4.2 The role of the economic actor

In addition to considering the role that the user can and should play within AI regulation, it is also necessary to discuss the role that those responsible for these AI systems should play within the regime. This may include those designing the AI, manufacturing physical components for an AI-based system, and those distributing, importing, or deploying the system itself; these are typically the economic actors responsible for ensuring the safety of their products once in the hands of the consumer.[627]

AI is slightly different from a traditional product in that by nature it is designed to continuously change its output post deployment (this is especially the case for more sophisticated AI's). Therefore, it is reasonable to suggest that AI should be treated differently to that of a traditional product via traditional product safety regulations for the very reason that it can be seen to have its own distinct life cycle.[628] Nonetheless, without placing specific burdens upon economic actors, we cannot ensure the development of 'safe' AI.[629]

Figure 11: AI Life Cycle



Common sense would suggest that the initial burden for complying with regulatory measures should be placed upon those designing the AI system, alongside those directly involved in the engineering and manufacturing of a system that encompasses an AI (e.g., an autonomous robotic arm within a factory). But there are other parties who should be equally targeted by the regulations including those distributing the system, for example retailers,

[627] Office for Product Safety and Standards, ' Guidance: Product safety advice for businesses' (gov.uk, 2021) <https://www.gov.uk/guidance/product-safety-advice-for-businesses> accessed 25/06/2022
[628] A. Circiumaru, 'Three proposals to strengthen the EU Artificial Intelligence Act' (2021) *Ada Lovelace Institute* < https://www.adalovelaceinstitute.org/blog/three-proposals-strengthen-eu-artificial-intelligence-act/#:~:text=The%20European%20Commission's%20Artificial%20Intelligence,in%20law%20enforcement%2C%20education%20and> accessed 25/06/2022
[629] Detail regarding the specific obligations to be placed on economic actors can be found further on in this Chapter.

importers of systems and even the authorised representatives of manufacturers based outside of the country of implementation. Again, this is an approach championed in the forthcoming PSTI Bill, and one that appears to adequately extend the scope of the regulations to cover the entire life cycle.[630]

As per the PSTI Bill, compliance with regulatory measures can usually be achieved in numerous ways, with the priority being to lessen the burden of compliance on the economic actors in question. This is where industry standards are of use; one can list a technical regulatory requirement clearly within a piece of legislation and also make available an interchangeable designated standard that may be complied with in its place – implementing either within the manufacturing process will result in compliance with the overall regulations.[631] Therefore, if an economic actor has already declared its compliance with a recognised provision of a designated standard, then there is no additional need for them to declare compliance with a newly introduced legislative provision.

This is a key consideration for the introduction of any new regulatory measures; if the measures are too onerous for an organisation or body to implement then compliance will remain low. Alternatively, if an effort is made early on to ensure that compliance with the new measures is relatively simple, e.g., by offering multiple routes to compliance, then overall compliance will be much higher from the outset. This approach can and should be applied within the regulation of AI, however, the issue we face at present is that there is a serious lack of global AI technical standards that we could designate within legislation.[632] We are beginning to see efforts to develop such standards, for example via the newly created UK AI Standards Hub,[633] although until the point at which we have these robust, universally recognised standards, utilising the concept of designated standards within the regulatory approach will not be a viable option.

6.5 The Command

The next core regulatory principle to be considered is the command, e.g., what do the regulations and rules command of the target. The command makes up the bulk of any regulatory regime and is therefore arguably the most integral part of the regime in its entirety. With regards to AI, there are several competing views as to what specifically we

[630] Ibid n622
[631] Ibid n622
[632] The Alan Turing Institute, 'New UK initiative to shape global standards for artificial intelligence' (turing.ac.uk, 2022) <https://www.turing.ac.uk/news/new-uk-initiative-shape-global-standards-artificial-intelligence> accessed 30/06/2022
[633] Department for Digital, Culture, Media and Sport, 'New UK initiative to shape global standards for artificial intelligence' (GOV.uk, 2022) <https://www.gov.uk/government/news/new-uk-initiative-to-shape-global-standards-for-artificial-intelligence> accessed 30/06/2022

should be asking of the regulatory targets in order to encourage safe use of AI. For example, the UK would prefer to impose rules dependent on sector,[634] and the US are also adopting a similar approach.[635] In contrast to this is the EU AI Act which is renowned for its blanket style approach.[636]

As discussed in Chapters Three and Four of this thesis, there are a number of benefits and shortcomings for each of these approaches, therefore drawing upon the concepts that work well within these proposals and identifying the universal disadvantages is necessary to establish a functional set of rules. This section considers three key principles that should be central to the regulatory framework proposed within this thesis, this includes an amendment to the risk-based approach recommended by the EU Commission, the inclusion of education enhancement measures, and creation and use of best practice models.

6.5.1 The Risk-Based Approach

Using a risk-based approach to categorise AI for the purposes of regulation is logical and is favoured by the EU, as evident in their proposed AI Act.[637] This approach has a number of merits; the general structure is easy to follow for those unfamiliar with AI and its associated risk factors, and it provides a degree of certainty for those developing AI as to what is required of them in order to achieve compliance.

Yet, as discussed within Chapter Four, the approach does have its shortcomings; most of the categories are ill-defined, and at its heart this approach does not achieve the human-centric aims that the overall Act claims to accomplish, meaning that it does not adequately protect fundamental rights.[638] This does not mean however that using a risk-based approach to regulate AI is obsolete, in fact, this author is of the opinion that using a risk-based method of categorisation is beneficial. However, this approach as it stands within the EU proposal is not fit for purpose, and therefore for this risk-based strategy to work some changes should be made.

Firstly, we must consider what exactly it is that we want to categorise. At present, the EU AI Act places AI applications into four categories based upon their potential risk of causing harm. There are a couple of issues with this; primarily, there isn't really any reference made to the actual harm that might be caused by these applications, and there appears to be

---

[634] Office for AI, Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, 'National AI Strategy' (2021) <https://www.gov.uk/government/publications/national-ai-strategy> accessed 23/11/2022

[635] Y. Chae, 'U.S. AI Regulation Guide: Legislative Overview and Practical Considerations' (2020) *The Journal of Robotics, Artificial Intelligence & Law* 3(1) 17-40, 17

[636] Ibid n8

[637] Ibid n8

[638] See Chapter 3 for further detail

potential for overlap of the categories. For example, the unacceptable risk category will include "all AI systems considered a clear threat to the safety, livelihoods and rights of people."[639] These applications will be banned, however, there are applications that could fall within this category that could quite easily fall within the high-risk category as well, so it becomes unclear as to what truly sets these categories apart.

Therefore, any attempt to devise a working risk-based approach for AI regulation must address this issue and ensure that there is clarity regarding which application should fall within each of the risk categories. Further to this, it must be clear from the outset which fundamental rights are protected by each of the risk categories (this is not the case within the current EU proposal). This is because the overall aim of the regulatory regime proposed within this thesis is to secure AI and protect and empower AI users. Therefore, a risk-based approach that focuses more specifically on the actual and potential harms likely to result from the use of AI applications would more effectively protect our fundamental rights.

An example of how the risk-based approach as it stands in the EU proposal may not work in practice can be found by considering the case of the seven year-old child injured by a chess-playing robot.[640] In this particular case the robot broke the child's finger during a chess match after the child apparently rushed the robot by moving too quickly on the board.[641] In response to the child's quick movement the robot arm grabbed the child's finger, keeping hold of it and squeezing it tightly until bystanders were able to free the young boy from its grip.[642] It is believed that the robot acted in such a way as it expects a certain amount of time between moves, and when the child did not adhere to this expectation, the robot reacted in an unexpected way.[643]

First of all, if we were to place such a robot (pre-incident) into the EU risk-based framework it would seem that the system might fall into the minimal or no risk category seeing as this category includes AI used in games.[644] The robot certainly does not seem to fall within the unacceptable risk category, as a chess-playing robot does not typically pose "a clear threat to the safety, livelihoods and rights of people".[645] It also does not appear to be one of the AI

[639] Ibid n518
[640] M. Angelova, M. McCluskey, 'Chess-playing robot breaks boy's finger at Moscow tournament' (edition.cnn.com, 2022) <https://edition.cnn.com/2022/07/25/europe/chess-robot-russia-boy-finger-intl-scli/index.html> accessed 25/07/2022
[641] Ibid
[642] Ibid
[643] S. Sharwood, 'Russian ChessBot breaks child opponent's finger' (theregister.com, 2022) <https://www.theregister.com/2022/07/25/russian_chessbot_breaks_players_finger/> accessed 25/07/2022
[644] Ibid n316
[645] Ibid n316

types listed within the high-risk category, as this list centres around AI that is typically used within the public sector, health care, recruitment, law enforcement etc.[646] The limited risk category appears to primarily concern itself with chatbots, and so therefore by default it appears to fall within the lowest risk category, meaning that under the proposed EU AI Act, the robot would not be made subject to any specific requirements.

Post-incident, the robot would likely fall within the unacceptable risk category as it has caused physical harm to a young child, but this would not be the case until the actual harm had already been caused. Therefore, this would render the present risk-based approach ineffective; if an AI can only be regulated or banned after it has caused harm in some way, then the regulations seem retrospective as opposed to preventative, and therefore dysfunctional. This example clearly demonstrates the shortcomings of the current risk-based strategy, and therefore gives grounds for a new approach to build upon this.

[646] Ibid n316

Figure 12: A Rights-based AI impact assessment[647]

The below risk assessment matrix contains some example entries which indicate what filling the form out for an AI application might look like in practice.

AI application under assessment in this example: An algorithm used by judges to help calculate appropriate prison sentences.[648]

Right Impacted	How the right has/will be impacted upon	Level of Impact (Low/Medium/ High)	Level of certainty regarding impact (Certain/some certainty/uncertain)	Are testing outcomes expected/reliable?	Has the system undergone any external assurance/compliance? (Yes/No, provide details if yes)	Any issues reported re: performance of system? (If yes, what, and how will these be addressed?	Risk Rating (Low/Medium/ High)
Justice (right to due process/fair trial)	Non-transparent algorithm being used to make decision that will determine prison sentence – may affect right to due process.	**High** - use of algorithm in this setting might prevent right from being exercised entirely.	Certain – there is a high likelihood that there will be challenges brought regarding the use of this algorithm and its conflict with this right.	Testing is mostly accurate although there is some evidence of inaccurate test results.	No	No	**High**
Equality (protection from discrimination)	Factors taken into consideration by the algorithm could be classed as protected characteristics.	**High** – using these factors may lead to discrimination (also by proxy).	Some certainty – it is somewhat likely that there will be challenges brought regarding the functioning of this algorithm and its conflict with this right.	Some outcomes have proven to be potentially influenced by biased data.	No	Some initial testing has shown potential evidenced of biased data being used to train the algorithm.	**High**
Overall Risk Rating and Category Low/Medium/High							**High**

[647] As per the European Convention on Human Rights, the Charter of Fundamental Rights of the European Union and the Universal Declaration of Human Rights

[648] This is a fictional example for the purposes of this thesis.

6.5.1.1 Rights-based risk assessment explained

The rights-based risk assessment for AI proposed here is designed to both calculate the risk category of any given AI application, how likely it is that the application will impact upon the various fundamental rights (examples of which are provided in an accompanying list), and identify what measures are necessary to be taken by the manufacturer in order to ensure the deployment of the safest system possible. The strongest asset of this approach is that it is applicable to all AI applications, meaning that this particular method could be utilised regardless of sector and within both blanket and sector specific regulatory regimes. It is worthwhile however explaining the purpose of each of the assessment criteria included here, and overall, how each of these criteria can lead to an accurate risk rating being calculated.

Right Impacted

Here the individual conducting the risk assessment would be expected to identify any rights that either have been or are likely to be impacted upon by their application. This list is guided by the European Convention on Human Rights, the Charter of Fundamental Rights of the European Union, and the Universal Declaration of Human Rights. These rights include:

- Equality: This includes protection from discrimination, and other related rights. Examples of AI that might infringe upon this right include algorithms used to decide who is and is not eligible to receive a bank loan, or access to a visa etc.
- Freedoms: This includes the right to private life, protection of personal data, freedom of expression, assembly/association, education, asylum. Examples of AI that might infringe upon this right include the use of algorithms online to censor what material is publicly available to view, the use of facial recognition technology in public spaces (particularly by the state) and algorithms used to determine grades within school systems.
- Justice: This includes right to due process, right to a fair trial, presumption of innocence, right to defence. Examples of AI that might infringe upon this right include algorithms used to determine reoffending rates and calculate prison sentences.
- Solidarity: This includes the right to fair and just working conditions, social security and assistance, health care, consumer protection. Examples of AI that might infringe upon this right include algorithms used diagnostically within health care settings, and AI used to determine who may access social welfare and state benefits.
- Citizens' rights: This includes the right to good administration, access to documents. Examples of AI that might infringe upon this right include algorithms used by public sector bodies in public services, e.g., transportation or even social scoring.

- Dignity: This includes the right to life, integrity, degradation, physical harm. Examples of AI that might infringe upon this right include autonomous weaponry, autonomous vehicles without adequate safety features included within its design, or systems with the capacity to cause physical harm such as a robotic arm etc.

How the right has/will be impacted upon

Here the individual responsible for conducting the assessment will provide information regarding how the right in question has or might be violated by the application, e.g., the data used to train the AI may contain biases, or there is little to no human oversight for the decisions being made and implemented by an algorithm.

Level of impact

This is one of the most important aspects of the risk assessment. Based upon the information provided by the individual in the earlier columns of the risk assessment table, those responsible for the AI are asked to consider the level of impact that their application might have on the particular right in question. This rating will help to determine the overall risk rating once the application is completed.

For example, a company has produced an algorithm that will help judges to determine prison sentence length, and they have identified via this risk assessment that their application might conflict with the justice right (namely the right to due process). They have identified this as possibility as they are protecting their algorithm as a trade secret, meaning that information regarding the functioning and reasoning of the algorithm would not be available for investigation by the court. They would then be asked to rate the level of impact on this right, there are three ratings to choose from: low, medium, and high. For the example given here, the algorithm would likely have a high impact upon the right to due process for the reasons stated above, as it would for the most part prevent that right from being exerted completely.

Level of certainty regarding impact

Here those responsible for the design of the AI are asked to consider how likely it is that their AI will impact upon the specific right in question; is it certain that it will, possible that it will, or are they uncertain that it will. Again, this particular question is posed to help to determine the overall risk rating of the application.

Testing outcomes

This question is especially important for more sophisticated AI with machine learning capabilities. Here manufacturers are asked to assess their training data and whether or not it has been ethically sourced/freed from bias, how the system has performed so far and

whether or not these tests have resulted in outcomes that were expected. This last point is particularly important as considering this data will give indication as to whether the system is performing as intended, or if it is likely that it will produce unintended results, outcomes and decisions as the chess-playing robot discussed earlier in this chapter did.

Ensuring that there is adequate testing, verifying, and retesting procedures within any AI life cycle is crucial as it means that developers are aware of the potential risks their system might pose, and that they are taking action to mitigate unintended results at an early stage. This is also a helpful indication for the overall risk rating of the device, as if there is if the system isn't performing well during testing, then the device will be rated as an overall higher risk system.

External assurance and compliance

This section of the assessment asks the manufacturer to provide information regarding whether or not their system has undergone any assurance testing, or whether their system is compliant with any industry standards. If a system has undergone assurance testing and is compliant with a recognised standard, then this would increase the reliability of the system and could lower the overall risk rating of the system.

Reporting results

The final part of the risk assessments asks those responsible for the device to note whether or not there have been any issues with the system reported. Earlier in this chapter it was proposed that there should be a mandatory requirement for those developing and deploying AI to have a clear and accessible reporting procedure made available via their websites for example. This empowers users to report any issues that they have found during their use of the system and would assist in determining the overall risk rating of the device.

Risk rating

Finally, making use of the information provided a risk rating can be calculated and the system will be rated as either high risk, medium risk, or low risk. When calculating this rating, both the level of impact and likelihood of impact are some of the most important factors, but the other information provided within the assessment will be useful in determining the rating also.

It is also worth noting that a system might be at risk of impacting upon not one but multiple rights, in such a case the overall risk categorisation of the system would be based upon an assessment of multiple risk ratings in total.

The overall rating will then indicate which measures are necessary for the system to be safely deployed. For example:

- High risk system:
    - It is necessary that adequate transparency by design principles are built into the system from the outset, meaning that the system is much more transparent in its function.
    - The system must be explainable, this means that algorithms within this category should not be treated as trade secrets (as per the example given earlier) as they should be accessible to those investigating their functions.
    - The system must be secure. Manufacturers must ensure that their systems contain adequate cybersecurity features as most of these systems will likely deal with highly sensitive personal information and this must be safeguarded.
    - Manufacturers and designers of systems within this category must show that they have adequate and consistent testing, review, and verification procedures in place. And they must provide evidence that they are conducting such reviews sufficiently frequently.
    - It is also necessary for these systems to retain a sufficient level of human oversight to minimise risk of harm from occurring.
    - It may be required that these systems are compliant with specific industry standards to improve robustness of the system (once these standards are developed).
    - A reporting system by which users can report system flaws is necessary.

- Medium risk system:
    - Transparency by design principles should be present and built into the system from the outset to ensure a level of transparency for the system (this can help with system longevity, improvement, and development).
    - There must be clear information communicated with the user of the system to indicate that the system they are using is autonomous and what type of information is being utilised by the system.
    - Manufacturers and designers of systems within this category must show that they have adequate and consistent testing, review, and verification procedures in place. And they must provide evidence that they are conducting such reviews sufficiently frequently.

    - In some cases, it may be required that algorithms within this category should not be treated as trade secrets, dependent on their use (again, in order to aid explainability).
    - A reporting system by which users can report system flaws is necessary.

- Low risk systems:
    - Transparency by design principles should be present and built into the system from the outset to ensure a level of transparency for the system (this can help with system longevity, improvement, and development).
    - Manufacturers and designers of systems within this category should promote best practice by showing that they have adequate and consistent testing, review, and verification procedures in place.
    - A reporting system by which users can report system flaws is necessary.

There are a number of ways in which these requirements could be mandated. The method with the highest likelihood of success would be introducing primary legislation that mandates implementation of these requirements and providing supporting documentation such as the list of example rights and applications provided earlier in this sub-section in a piece of amendable secondary legislation.

Second to this would be to introduce these measures via a code of practice or similar document. This would mean that the requirements listed here would not be legally binding, however they could indicate industry best practice which may incentivise uptake. The first of the two options would be preferable, but as discussed within Chapter Five it is recognised that developing legislation is time consuming, and so in order to see quick implementation, drafting a code of conduct detailing these requirements may be the most logical initial step.

Benefits of this approach

There are a number of reasons why this approach would be beneficial in comparison to those already suggested within various national AI strategies. First and foremost, this approach is truly human-centric; this means that all those involved in the development, deployment and management of AI would have to directly consider the impact that their system may have upon a selection of fundamental rights. Right from the outset this means that the likelihood of rights violations will be lower as opportunity would be provided throughout the life cycle of the AI for identification of fundamental rights conflicts.

Furthermore, if we compare with this approach to that suggested by the EU for example, it would appear that the scope of the proposed risk assessment framework here is wider

reaching and the risk categories are better defined as a result. If we consider assessing the earlier example of the chess-playing robot that injured the young Russian child against the rights-based risk assessment proposed here, the bot would likely be identified as potentially posing a physical risk much earlier than if the EU risk-based approach was applied. This would therefore mean that the potential flaws in the robot would be identified at a much earlier stage, which would then provide opportunity for earlier intervention and harm mitigation.

It is also worth noting the reasoning behind the structure of the proposed risk assessment framework proposed here. By categorising the AI based upon the specific rights that it may infringe upon, and then categorising by risk, we are allowing for a more flexible mode of regulation. For some of the EU risk categories, there are rather exhaustive and limiting lists of AI applications that would fit within each of the categories. As per Chapter Five of this thesis, ensuring flexibility and futureproofing is of utmost importance when regulating modern technology.[649] Therefore, rather than describing outright the types of AI that might violate each of the fundamental rights listed, it would be preferable to include these within a separate list (this list could be included within secondary legislation for example or accompanying guidance which would make the list more easily amendable and therefore flexible to change).

Requirements dependent on risk rating - Implementing the key AI principles

In Chapter Five of this thesis, transparency was discussed as the key ethical principle that needs to be included within regulation on AI. There are several actions that economic actors should take (dependent on risk rating) in order to implement transparency, which in turn would allow them to mitigate the risk of their systems violating fundamental rights and ensure compliance with the governance framework proposed within this chapter.

One notable requirement is for developers of AI to implement transparency by design principles from inception of the system.[650] This would mean that transparency as a concept would be embedded within the system as opposed to be tacked onto the system as an afterthought, most likely after an incident has already occurred (similar to the approach championed by GDPR via privacy by design).[651]

[649] E. Jackson, Regulating Reproduction: Law, Technology and Autonomy (Bloomsbury, 2001)
[650] H. Felzmann, E. Fosch-Villaronga, C. Lutz, A. Tamo-Larrieux, 'Towards Transparency by Design for Artificial Intelligence' (2020) Science and Engineering Ethics 26, 3333-3361
[651] Privacy Policies, 'Implementing Privacy by Design' (privacypolicies.com, 5 January 2021) <https://www.privacypolicies.com/blog/privacy-by-design/#What_Is_Privacy_By_Design> accessed 01/08/2022

Systems with intelligent capabilities should therefore be designed in such a way that safety and security are made to be paramount features, and these systems should be as user-friendly as possible (namely, they should not be built in such a way that exploits users).[652] The rights-based impact assessment proposed here would go some way to ensure that transparency by design is achieved, as again, from the outset the manufacturer is required to consider the potential impact their device might have upon various fundamental rights. The higher the risk the system is deemed to have of violating fundamental rights following the impact assessment, the more robust the transparency by design features implemented into that device may be. Codes of practice and general guides might help here in guiding manufacturers as to what types of measures they can implement in order to achieve transparency by design.

In a similar vein, explainability may also be achieved via this method (another requirement dependent on risk rating). By making a system more transparent the manufacturer is by default making the system easier to understand, however, there is more that we can do to ensure explainability is embedded within AI; by improving technical education (as discussed within in Chapter Five).[653] The burden for implementation of this would likely be borne by the state as opposed to the manufacturer, yet it goes without saying that improving the technical understanding of the public, and anyone who has regular contact with AI-based systems would go far to improve explainability in general.

Declaration of Conformity

It is also necessary to request that manufacturers (and relevant economic actors) make a formal, public declaration that their system has undergone the aforementioned impact assessment and that they have implemented the necessary requirements as per their systems risk rating. Typically, this type of declaration is used to provide evidence that a product has conformed with particular standards prior to being placed on the market.[654]

The declaration of conformity offers a level of reassurance to those using the product, device, or system, and promotes trustworthiness (a key goal for AI). This declaration may be contained within the organisations website, in documentation that accompanies the product or system, or via a pop-up notice when the system is in operation. Therefore, by requiring manufacturers to make available a clear and concise declaration that their system has

[652] M. Boden, J. Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, P. Newman, V. Parry, G. Pegmanz, T. Rodden, T. Sorrell, M. Wallis, B. Whitby, A. Winfield, 'Principles of Robotics: regulating robotics in the real world' (2017) *Connection Science* 29(2) 124-129
[653] S. Reddy, S. Allan, S. Coghlan, P. Cooper, 'A governance model for the application of AI in health care' (2019) *Journal of the American Medical Informatics Association* 27(3) 491-497
[654] Health and Safety Executive for Northern Ireland, 'Declaration of Conformity' (hseni.gov.uk, 2022) < https://www.hseni.gov.uk/articles/declaration-conformity> accessed 01/08/2022

undergone the proposed risk assessment, it is likely that there will be increased trust amongst those using the system.

6.6 The Consequences

Now that we have established the potential regulator, the target audience, and the specific requirements of the proposed regulations, we must consider the final regulatory principle; the consequences.[655] Without consequences, the command, or rules we are imposing will become obsolete; why would a manufacturer implement such rules if there is no risk of repercussion?

It is important to note however that despite the common understanding that consequences are usually negative, this is a misconception; consequences may also be positive.[656] A prime example of a negative regulatory consequence is a fine such as those imposed via GDPR for non-compliance.[657] Whether or not this particular consequence actually functions effectively as a deterrent is questionable,[658] yet the use of fines in this context is commonplace.

6.6.1 Punitive Penalties

As per the above, fines have been used as sanctions within data protection regimes for some time.[659] Similarly, fines are also commonly used within competition law to strengthen enforcement and encourage compliance.[660] In summary, they are typically used to deter non-compliance and strengthen the general enforcement of a regulatory regime. In addition, it is believed that the fine should be representative of the damage caused; for example, a data breach within a well-established organisation that directly impacts upon a large portion of the population should warrant a larger fine than a data breach that occurs on a smaller scale and effects less people.[661]

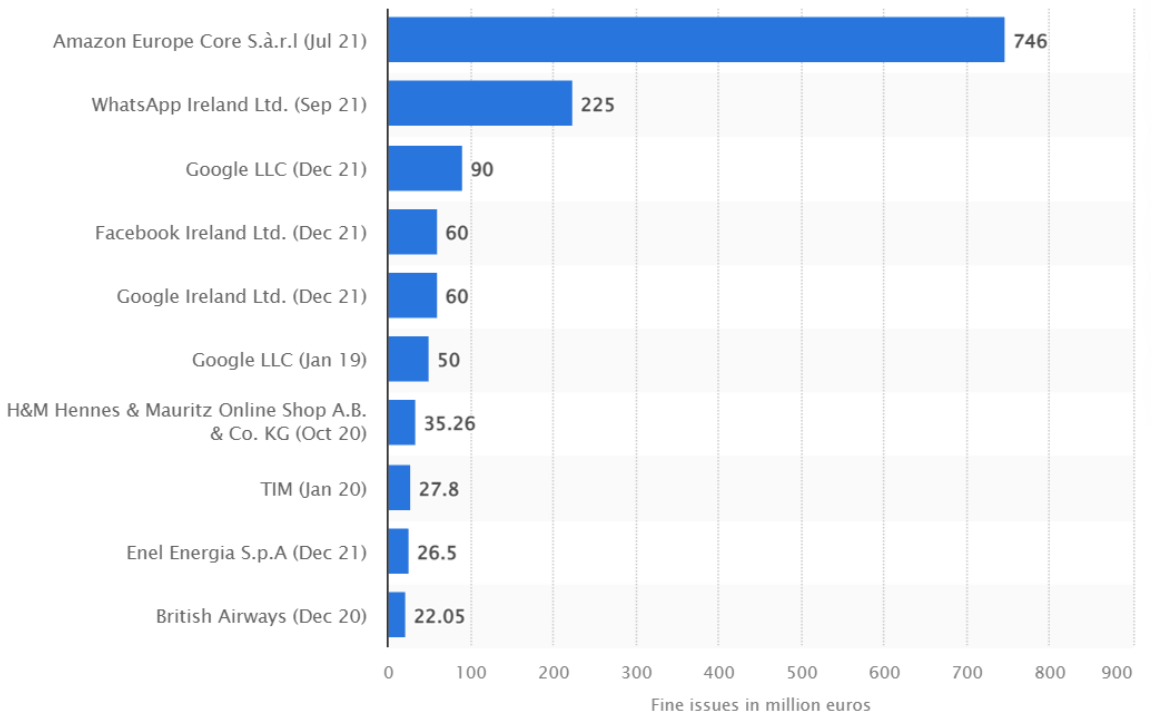
In general, the punitive punishments that GDPR has introduced have had wide-ranging effect, from generating an entire GDPR consultancy market,[662] to triggering some organisations to delete in entirety their customer records (including email addresses etc) as

---

[655] Ibid n584
[656] Ibid n584
[657] Ibid n25
[658] M. N. Lintvedt, 'Putting a price on data protection infringement' (2021) *International Data Privacy Law* 12(1)
[659] Ibid
[660] European Commission, 'Fines for breaking EU competition law' (ec.europa.eu) <https://ec.europa.eu/competition/cartels/overview/factsheet_fines_en.pdf> accessed 01/08/2022
[661] W. P. J. Wils, 'E.C. Competition Fines: To Deter or Not to Deter' (1995) *Yearbook of European Law* 15(1) 17
[662] D. Allen, A. Berg, C. Berg, B. Markey-Towler, J. Potts, 'Some Economic Consequences of the GDPR' (2019) *Economics Bulletin* 39(2) 785-797

opposed to having to actively comply with the regulations.[663] Interestingly though, whether or not public bodies are subject to fines for non-compliance differs state by state, as does the calculation of fines.[664]

Figure 13: Largest fines for GDPR violations (accurate as of Summer 2022)[665]



Source: Statista, 'Largest fines issued for General Data Protection Regulation (GDPR) violations as of July 2022' (2022) https://www.statista.com/statistics/1133337/largest-fines-issued-gdpr/, accessed 01/08/2022

To date, we've seen several considerably large fines handed out to some of the world's biggest tech companies for non-compliance with GDPR, as per the figure above. The largest of these by far is Amazon's fine of €746 million handed to them for having a non-compliant cookie consent policy.[666] There are a number of messages that we can take from this information that may help in determining how fines could be used to enforce the regulation of AI. For example, Amazon had an annual revenue of $469.822 billion in 2021,[667] and minus the GDPR fine they were given that same year, Amazon still had a revenue of $468.244 billion. Therefore, the fine that they were given, despite being the largest GDPR fine ever

[663] Ibid n660
[664] Ibid n660
[665] Statista, 'Largest fines issued for General Data Protection Regulation (GDPR) violations as of July 2022' (statista.com, 2022) <https://www.statista.com/statistics/1133337/largest-fines-issued-gdpr/> accessed 01/08/2022
[666] Tessian, '30 Biggest GDPR Fines So Far (2020, 2021, 2022)' (tessian.com, 2022) <https://www.tessian.com/blog/biggest-gdpr-fines-2020/> accessed 01/08/2022
[667] Statista, 'Annual net sales revenue of Amazon from 2004-2021' (statista.com, 2022) https://www.statista.com/statistics/266282/annual-net-revenue-of-amazoncom/ accessed 01/08/2022

given, was a mere ‘drop in the water’ in terms of their annual revenue (which continues to grow year on year).

Compare this to a company with a much smaller annual turnover; whilst GDPR does take into consideration an organisation’s annual profits when calculating the appropriate fine to hand out, a large fine given to a much smaller company has a higher likelihood of either bankrupting the company or at best deterring them from a repeat violation. Whereas for companies like Amazon, Facebook (Meta), or Google, their annual revenues are so high that it would be hard for a GDPR fine to ever have a significant impact upon it, and therefore there is a lesser chance that they would see it as a deterrent.

There is also the public image consideration which these companies will be conscious of, e.g., will the public continue to trust a company with their data if the same company continues to breach data protection regulations. However, as companies such as Amazon, Facebook and Google continue to saturate the market, it becomes increasingly more difficult for the general public to avoid using their services.

Using GDPR enforcement as a lesson here, it would appear that using fines as a means to enforce compliance with regulations has some merit; if the fines are large enough (and proportionate to the harm caused) they should act as a deterrent. However, as demonstrated, it is questionable as to whether these fines actually act as a deterrent when the company is large and influential enough. This is a similar problem we may face in the regulation of AI; it is a similar group of companies that lead in AI development, such as Amazon, Google, and Microsoft.[668]

In a study on GDPR penalties conducted by Wolff and Atallah[669], it was revealed that there were several repeat offenders; those who had been identified as having committed separate violations for each of the fines. This suggests that whilst using fines in this capacity, despite the unprecedented and uncapped potential value of such fines, we are still seeing organisations relatively undeterred and continuing to violate regulations.

Therefore, we must be conscious that any punitive punishment scheme takes this into consideration. If it does not, we run the risk of adversely punishing smaller start-up companies developing AI, even rendering them bankrupt for non-compliance, whilst allowing the large tech companies in this space to continue relatively unphased and somewhat unaffected by the fine they’ve been given.

---

[668] J. Maguire, ‘Top Performing Artificial Intelligence Companies of 2022’ (datamation.com, 2022) https://www.datamation.com/featured/ai-companies/ accessed 01/08/2022

[669] J. Wolff, N. Atallah, ‘Early GDPR Penalties: Analysis of Implementation and Fines Through May 2020’ (2021) *Journal of Information Policy* 11 63-103

Applying what we know about the GDPR fines scheme it would be reasonable to suggest that in terms of regulating AI:

- Fines should also be given to public sector bodies who do not comply with the proposed regulations, in the same way they are given to private companies.
- A fine scheme should be standardised across borders, meaning that when awarding fines there is a general guideline that is followed in order to promote international harmonisation.
- Individuals that have suffered as a result of non-compliance should be compensated. This is suggested by Lintvedt et al as a potential solution for improving GDPR,[670] but this suggestion would also work well alongside the regulations proposed within this thesis in that it would go towards ensuring the regulations are more human-centric.

6.6.2 Other methods of enforcement

Considering the enforcement of product safety regulations is also useful when determining the most effective way to ensure compliance with future AI regulations. In particular, the use of market surveillance may be quite useful. With regards to product safety, local authorities are tasked with conducting market surveillance by examining and testing products to ensure that they are safe, and if necessary, conduct a further investigation.[671] The regulator (as discussed earlier in this chapter) is also equipped with similar investigatory powers, which would also be suitable for the regulation of AI.[672]

Market surveillance is particularly beneficial in this sense as it allows for any potential non-compliance to detected at an early opportunity prior to harm being caused, therefore mitigating potential costs and offering organisations live feedback on the functionality of the systems that they have deployed.[673] This goes hand in hand with the proposed requirement for organisations to have in place an adequate vulnerability reporting procedure, in that organisations would have several opportunities for receiving feedback on their systems before any large scale harm is actually caused. This therefore means that via these measures we could effectively reduce the rate at which harm is caused by AI, and in some cases prevent it entirely.

---

[670] Ibid n658
[671] Ibid n584
[672] Ibid n584
[673] Exactpro, 'Exactpro Test and Automation Approaches: A Case Study in Market Surveillance' (exactpro.com, 2022) < https://exactpro.com/case-study/market-surveillance#:~:text=The%20optimisation%20of%20the%20Market,requirements%20are%20met%20and%20documented.> accessed 01/08/2022

There would be cost implications in requiring local authorities and a potential new regulator to develop testing suites as well as train individuals to test these systems and carry out the market surveillance. However, this could be outsourced to existing testing houses that have been accredited by the regulator, or relevant standards body. For example, UKAS (UK Accreditation Service) offer accreditation to biological laboratory testing facilities within the UK if they are found to be compliant with ISO/IEC standard 17025.[674] Therefore, as opposed to focusing on funding new testing facilities for the regulator and local authority, we could focus on accrediting established testing facilities, which would overall reduce any initial financial implications.

6.7 Conclusions

Overall, this Chapter proposes a regulatory framework for the regulation of AI. The framework aims to ensure a well-rounded regulatory regime is in place that balances the need to secure, protect and empower AI users, whilst ensuring continued investment, development, and innovation in the AI sphere. To achieve this, the framework relies upon the adoption of a rights-based AI impact assessment; this impact assessment allows organisations to identify any potential fundamental rights violations that their systems might cause and gives ample opportunity for these organisations to receive continual feedback throughout the life cycle of their systems as to the functionality of their systems.

This framework is formed by considering the various regulatory approaches proposed by the nations, regions and international bodies examined within Chapters Three and Four of this thesis. The framework is such that it could be adopted by any number of states, international bodies and organisations and amended to fit specific constitutional and administrative models (although the proposed framework does refer to UK infrastructure e.g., when discussing regulators merely as an example). This Chapter therefore achieves its goal in addressing the fourth and final research question set out within this thesis; what realistic and workable recommendations can be made to improve and secure the current state of AI regulation?

The framework proposed within this chapter is only a starting point, but one that addresses the shortcomings identified in a number of regulatory methods currently proposed by a variety of states and so achieves several aims that these do not, e.g., this framework is human-centric whereas the one proposed by the EU is not. By introducing regulatory measures such as this, we can truly begin to build a regime that fully supports the

[674] UKAS, 'Laboratory Accreditation – Biological' (UKAS.com, 2022) < https://www.ukas.com/accreditation/standards/laboratory-accreditation/biological/> accessed 01/08/2022

development of safe and trustworthy AI, whilst placing the end-user at the heart of the regime.

Conclusion

This thesis establishes several issues intrinsic to the relationship that exists between AI and the law; there is widespread and relatively unchecked use of AI, the occurrence of unintended consequences as a result of this use is growing, and the current state of AI regulation is uncertain and susceptible to considerable weaknesses. The thesis begins by outlining these three issues as central research questions that have guided the work undertaken throughout this project and have such formed a basis for the overall contributions of this research.

It was identified in the introduction to this thesis that whilst the literature on the topic of AI regulation is growing, it is limited in its impact; that is to say that whilst many scholars identify weaknesses in the law and issues surrounding AI use, very few (if any) offer meaningful suggestion with regards to how we could tackle the issues discussed. Therefore, we have an increasing amount of literature that warns us of associated AI risks, yet we are left feeling uncertain as to how we actually fix these issues. Therefore, the work presented within this thesis contributes to this gap in the literature by expanding upon existing knowledge regarding potential AI risks and offers a solution that will encourage the deployment of safe and reliable AI. This proposal is made with the interconnected and truly borderless nature of modern technology in mind.

The first chapter of this thesis introduces us to the existing relationship between AI and the law, defining AI for the purposes of this research, and using use case examples to demonstrate both the benefits and drawbacks of the technology. Defining AI is a task in and of itself, something that is considered in detail during this initial chapter. A definition is established however and is one that draws upon the multidisciplinary nature of AI by referring to it as an umbrella term for numerous types of technology. Referring to AI in such a way has several benefits, most importantly this broad definition allows for flexibility in an ever-changing technical landscape, whereas a more precise definition would be too ridged and wouldn't allow for the same freedom of scope. The examples used within this chapter lay the foundations for the discussions that follow in this thesis, by establishing early on the strained relationship that exists at present in many aspects between AI and the law.

Following this, the second chapter of this thesis goes further in exploring this relationship by presenting an in-depth analysis of the risk of discrimination and bias within AI-based systems and uses this as a case study to call for better regulation. This chapter, published in 2021, considers a number of recent instances of bias and discrimination that have occurred

in differing fields of AI application such as within the public sector, the financial sector and more generally amongst some of the biggest tech companies of our time including Google and Microsoft. Against this backdrop, it is established that most of the legal instruments we have in place at present are not adequately equipped to tackle this issue and as a result, are in need of effective reform, and in some places considerable overhaul. To this effect, a number of suggestions are made for the strengthening of these measures, and overall, this analysis serves as a basis for the larger scale proposals to follow in later chapters.

In an effort to assess global understanding of the legal vacuum that exists with regards to AI, the current state of AI regulation was assessed. This analysis resulted in a number of findings; firstly, there does appear to be a global understanding that AI associated risks must be addressed via regulation, however agreement as to the type of regulation is scarcely agreed upon. Secondly, there appears to be general acknowledgment that the overall aim of any such regulation should be to increase AI trustworthiness. And thirdly, AI is and will continue to be a large factor in most nations' economies in the coming years and so therefore continuing to promote innovation in the technology itself is essential. As a result, we are left regulatory proposals that differ widely in their approach (e.g., the UK and US appear to favour a sector-led regulatory approach whereas the EU favours a blanket-style legislation).

The regulatory proposals we have at present have numerous strengths (such as the EU's risk-based approach), but they also have considerable weaknesses (e.g., the risk categories within the EU proposal seem to be too many in number, and rather vague). This thesis offers a number of reasonable suggestions to amend these weaknesses and uses both the strengths and weaknesses discussed in Chapters Three and Four to promote a new AI governance framework in Chapter Six.

An analysis of the principles and key concepts necessary for any regulatory regime that targets a type of technology follows this and raises several issues that must be addressed within a functioning regulatory approach such as the need to consider whether or not legislation is a necessary tool, and the role that technical industry standards should play in the space. Chapter Five also evaluates the five paradoxes as presented initially by the Honourable Justice Michael Kirby, which are five key issues regarding the regulation of modern technology. Using these five paradoxes as a basis, a number of key concepts and principles that need to be embedded within any functioning AI governance model are presented, these are: transparency, explainability and security. A number of ways in which these principles can be achieved are discussed, including promoting the inclusion of transparency by design principles during the manufacturing and design process, improving

education and skills-based learning for AI, and ensuring industry standards are kept with regard to the cybersecurity of AI systems (particularly those processing large quantities of personal data).

The suggestions made here form part of the proposal that follows in Chapter Six, specifically the rights-based risk assessment framework recommended within this section. This proposed framework contains guidance notes that reference how these key principles recommended within Chapter Five, would be embedded, and tested within this framework. Further to the proposal made here, a number of suggestions are made regarding other crucial regulatory elements, such as who the relevant regulating body would be in this instance, who the ideal target audience should be, and what penalties should be in place for non-compliance with regulations. This thesis therefore contributes a proposal where the literature is lacking and provides foundations for further work on the regulation of AI to be completed.

Limitations and future research

There are limitations to any thesis, and there are certain factors that have constrained the research present here. A notable limiting factor that has presented itself throughout the course of this thesis has been the nature of the subject itself, i.e., the dynamic, changeable, and uncertain nature of AI. AI is developing at such an unprecedented rate that the literature and scholarly discussions on the topic itself are often difficult to keep pace with. For example, during the time it took to complete this thesis there have been several legislative proposals and case examples of AI harm that have considerably affected the direction and findings of this research. Therefore, new findings and legislative developments in this space has meant that sections of this thesis have had to be revisited several times.

Time has been another notable limitation within this thesis. Examining and proposing suggestions for the regulation of AI, is a project that is so large in scope that it will require further time offered to investigate it fully. This thesis provides a solid basis for this research to be continued, but due to thesis requirements and time-limitations, the research conducted here can only consider the regulation of AI to an extent.

Further research in this space is necessary and might consider questions such as:

- What are the most effective methods for future-proofing regulations on modern technology? This something that has been acknowledged as necessary within this thesis but as we see AI regulation come to life, only then will we be able to truly gauge how successful future-proofing methods have been and will continue to be.

- How do we regulate AI as it becomes more and more sophisticated? AI is ever-evolving, and in a similar-vein to futureproofing, we may need to rethink approaches to dealing with accountability and liability within legislative regimes
- What support will be necessary to support small and medium-sized enterprises (SMEs) in successfully adopting the regulatory measures? SMEs are often at the forefront of AI development and adoption, and it is integral that these businesses are able to continue to develop and deploy these systems. Therefore, we need to consider how we are going to support these businesses as new regulatory measures are adopted that will likely place onus on these types of organisations.

Overall, this thesis has evidenced that there is at present a strained relationship between AI and the law, exacerbated by the fact that there is a lack in effective AI regulation. There are some ideas regarding how we might regulate AI, but on the whole these proposals are flawed, and at present show significant weakness. Subsequently, this thesis establishes that AI governance is necessary, but that it must be approached in a careful way, thus demonstrating the balancing act of regulating in the age of AI and providing a basis for further research in the space to be conducted.

Bibliography

Books and Book Chapters

Blair, L. Choosing a Methodology in Writing a Graduate Thesis or Dissertation (Sense Publishers, Rotterdam 2016

Brown, A. C. Stern, J. Tenenbaum, B. Gencer, D. Handbook for Evaluating Infrastructure Regulatory Systems, The World Bank (2006)

Damore, D. F. Lang, R. E. Danielsen, K. A. 'Blue Metro's, Red States: The Shifting Urban-Rural Divide in America's Swing States' (Brookings Institution Press, Washington D.C. 2020)

Drożdż, A., Protection of Natural Persons with Regard to Automated Individual Decision-Making in the GDPR. (Kluwer Law International B.V., Netherlands, 2020)

Girasa, R. 'AI U.S. Policies and Regulations' in Artificial Intelligence as a Disruptive Technology (Palgrave Macmillan, London 2020)

Glass, R. Callahan, S. The big-data driven business: how to use big data to win customers, beat competitors, and boost profits (Wiley, New Jersey, 2015)

Glassman R. M., 'Will Artificial Intelligence (AI) Make Democracy Irrelevant?' in The Future of Democracy (Springer, 2019) 189-198

Graham, J. 'Risk of discrimination in AI systems: Evaluating the effectiveness of current legal safeguards in tackling algorithmic discrimination' in Lui, A. Ryder, N. (eds) FinTech, Artificial Intelligence and the Law: Regulation and Crime Prevention (Routledge, 2021)

Hoffmann-Riem, W. 'Artificial Intelligence as a Challenge for Law and Regulation' in Wischmeyer, T. and Rademacher, T.(eds), Regulating Artificial Intelligence (Springer, 2020)

Jackson, E. Regulating Reproduction: Law, Technology and Autonomy (Bloomsbury, 2001)

Kaiser, B. Targeted: My Inside Story of Cambridge Analytica and how Trump, Brexit and Facebook Broke Democracy (Harper Collins, 2019)

Katz, Y. Artificial Whiteness: Politics and ideology in artificial intelligence (Columbia University Press, 2020)

Kirk, J. Miller, M. L. 'Reliability and Validity in Qualitative Research' (SAGE Publications, 1986)

Lee, K. F. AI Superpowers: China, Silicon Valley and the New World Order (Houghton Mifflin Harcourt, New York, 2018)

McConville, M. Chui, W. H. Research Methods in Law (Edinburgh University Press, Edinburgh 2007)

Nilsson, N. J. Artificial Intelligence: A New Synthesis (Morgan Kauffman Publishers, San Francisco 1998) 1

Puntambekar, A.A. (2010) Design & Analysis of Algorithms (First). Pune: Technical Publication

Ranchordas, S. Van-t Schip, M. 'Future-Proofing Legislation for the Digital Age' in Ranchordas, S. Roznai Y. (eds) Time, Law and Change (Hart, 2020)

Schwab, K. (2017) The Fourth Industrial Revolution. London: Penguin

Simon, P. The visual organization: data visualization, big data and the quest for better decisions (Wiley, New Jersey 2014) 1
Van Zwanenberg, P. Ely, A. Smith, A. Regulating Technology: International Harmonization and Local Realities (Routledge, 2013)

Weaver, J. F. 'Regulation of Artificial Intelligence in the United States' in W. Barfield, U. Pagallo (eds) Research Handbook on the Law of Artficial Intelligence (Elgar, 2018)

Webley, L. 'Qualitative Approaches to Empirical Legal Research' in P. Cane and H. Kritzer (eds) Oxford Handbook of Empirical Legal Research (Oxford University Press, 2010)

Journal Articles and Conference Proceedings

Al Amaren, E. M. Hamad, A. M. A. Al Mashhour, O. F. Al Mashni, M. I. 'An Introduction to the Legal Research Method: To Clear the Blurred Image on How Students Understand the Method of Legal Science Research' (2020) International Journal of Multidisciplinary Sciences and Advanced Technology 1(9) 50-55

Allen, D. Berg, A. Berg, C. Markey-Towler, B. Potts, J. 'Some Economic Consequences of the GDPR' (2019) Economics Bulletin 39(2) 785-797

Ananny, M. Crawford, K. 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability' (2016) New Media and Society <https://journals.sagepub.com/doi/10.1177/1461444816676645>

Aruleba, K. Jere, N. 'Exploring digital transforming challenges in rural areas of South Africa through a systematic review of empirical studies' (2022) Scientific African 16 <https://doi.org/10.1016/j.sciaf.2022.e01190>

Baeza-Yates, R. 'Bias on the Web' (2018) Communications of the ACM 61(6) 54-61

Barocas, S. Selbst, A.D. (2016) Big Data's Disparate Impact. California Law Review. 104(3), 671–732

Bertot, J. C. 'Social Media, Open Platforms, and Democracy: Transparency Enabler, Slayer of Democracy, Both?' Proceedings of the 52nd Hawaii International Conference on System Sciences 2019 <https://scholarspace.manoa.hawaii.edu/bitstream/10125/61631/0782.pdf>

Boden, M. Bryson, J. Caldwell, D. Dautenhahn, K. Edwards, L. Kember, S. Newman, P. Parry, V. Pegmanz, G. Rodden, T. Sorrell, T, Wallis, M. Whitby, B. Winfield, A. 'Principles of Robotics: regulating robotics in the real world' (2017) Connection Science 29(2) 124-129

Borgesius, F.J.Z. 'Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence Algorithms and Artificial Intelligence' (2020) The International Journal of Human Rights. <https://www.tandfonline.com/doi/full/10.1080/13642987.2020.1743976?scroll=top&needAccess=true>

Boyte, H. C. 'John Dewey and Citizen Politics: How Democracy Can Survive Artificial Intelligence and the Credo of Efficiency' (2017) Purdue University Press Education & Culture 33(2) 13-48

Bruckner, M. A. 'The Promise and Perils of Algorithmic Lenders – Use of Big Data' (2018) Chicago-Kent Law Review 93(1) 1-59

Buiten, M. C. 'Towards Intelligent Regulation of Artificial Intelligence' (2019) European Journal of Risk Regulation 10 41-59

Buolamwini, J. Gebru, T. (2018) Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research. 81, 1–15

Burrell, J. (2016) How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms. Big Data & Society. 1–12

Cath, C. S. Wachter, B. Mittelstadt, M. Taddeo, L. Floridi, ‘Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach’ (2018) Science and Engineering Ethics 24 505-528

Calo, R. ‘Open robotics’, (2011) Maryland Law Review 70(3) 101-142

Cath, C. ‘Governing artificial intelligence: ethical, legal and technical opportunities and challenges’ (2018), Philosophical Transactions, Royal Society Publishing <https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2018.0080>

Chae, Y. ‘U.S. AI Regulation Guide: Legislative Overview and Practical Considerations’ (2020) The Journal of Robotics, Artificial Intelligence & Law 3(1) 17-40

Clifford Law Offices, ‘Driverless Car Accidents – Who is at Fault?’ (2021) National Law Review 11(176) https://www.natlawreview.com/article/driverless-car-accidents-who-s-fault

Coeckelbergh, M. ‘AI for climate: freedom, justice, and other ethical and political challenges’ (2021) AI & Ethics 1, 67-72

Coglianese, C. ‘Regulation’s Four Core Components’ (2012) The Regulatory Review <https://www.theregreview.org/2012/09/17/regulations-four-core-components/>

Cohn Adulamy, M. Shalev, V. ‘Colonscore: The Use of Machine Learning of Big Data to Detect Colorectal Cancer’ (2018) Harefuah (Israel Medical Association) 157(10) 634

Crowe, S. Cresswell, K. Robertson, A. Huby, G. Avery, A. Sheikh, A. ‘The case study approach’ (2011) BMC Medical Research Methodology 11 <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-11-100>

Cybenko, A. K. Cybenko, G. ‘AI and Fake News’ (2018) IEEE Intelligent Systems 33(5) 1-5

Doleac, J. L. Stevenson. M. T., ‘Algorithmic Risk Assessment in the Hands of Humans’ (ssrn.com, 2019) <https://ssrn.com/abstract=3489440>

Donnelly, D. ‘First Do No Harm: Legal Principles Regulating the Future of Artificial Intelligence in Health Care in South Africa’ (2022) Potchefstroom Electronic Law Journal (PELJ) 25(1) 1-43 < https://dx.doi.org/10.17159/1727-3781/2022/v25i0a11118>

Dressel, J. Farid, H. ‘The accuracy, fairness, and limits of predicting recidivism’ (2018) Science Advances 4(1) <https://advances.sciencemag.org/content/4/1/eaao5580>

Ehsan, U. Liao, Q. V. Muller, M. Riedl, M. O. Weisz, J. D. ‘Expanding Explainability: Towards Social Transparency in AI systems’ (CHI Conference on Human Factors in Computing Systems, May 2021) < https://arxiv.org/pdf/2101.04719.pdf>

Erdelyi, O. J, Goldsmith, J. ‘Regulating Artificial Intelligence: Proposal for a Global Solution’ (2018) AIES ‘18: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, February 2–3, 2018, New Orleans, LA, USA, 95-101, 96

Felzmann, H. Fosch-Villaronga, E. Lutz, C. Tamo-Larrieux, A. ‘Towards Transparency by Design for Artificial Intelligence’ (2020) Science and Engineering Ethics 26, 3333-3361

Felzmann, H. Villaronga, E.F. Lutz, C. Tamo-Larrieux, A. (2019) Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns. Big Data & Society. 1–14

Finck, M., 'Automated Decision-Making and Administrative Law' (2020) Max Planck Institute for Innovation & Competition Research Paper No. 19-10 https://ssrn.com/abstract=3433684

Floridi, L. Cowls, J. Beltrametti, M. Chatila, R. Chazerand, P. Dignum, V. Luetge, C. Madelin, R. Pagallo, U. Rossi, F. Schafer, B. Valcke, P. Vayena, E. 'AI4People – An Ethical Framework for a Good Society: Opportunities, Risks, Principles and Recommendations' (2018) Minds and Machines 28(4) 689-707

Fourie, L. Bennet, T. K. 'Super intelligent financial services' (2019) Journal of Payments Strategy & Systems 13(2) 151-164

Frankenberg, G. 'Critical Comparisons: Rethinking Comparative Law' (1985) Harvard International Law Journal 26(2) 439

Fredriksson, C. 'Big data creating new knowledge as support in decision-making: practical examples of big data use and consequences of using big data as decision support' (2018) Journal of Decision Systems 27(1) 1-19

Friedman, B. Nissenbaum, H. (1996) Bias in Computer Systems. ACM Transactions on Information Systems. 4(13), 330–347.

Gabriel, I. 'Artificial Intelligence, Values and Alignment' (2020) Minds and Machines, 30, 411-437

Gender Shades, 'Overview' (gendershades.org, 2018) http://gendershades.org/overview.html

Gesser, A. Maddox, R. Gressel, A. Colleluori, F. Lockwood, T. Pizzi, M. 'The EU AI Liability Directive Will Change Artificial Intelligence Legal Risks' (debevoisedatablog.com, 2022) < https://www.debevoisedatablog.com/2022/10/24/eu-ai-liability-directive/>

van Gestel, R. Micklitz, H-W. 'Revitilising doctrinal legal research in Europe: What about methodology?' (2011) European University Institute Working Papers < https://cadmus.eui.eu/bitstream/handle/1814/16825/LAW_2011_05.pdf?sequence=1&isAllowed=y>

Ghandi, S. Mosleh, W. Shen, J. Chow, C. M. 'Automation, machine learning and artificial intelligence in echocardiography: A brave new world' (2018) Echocardiography 35(9) 1402-1419 1402

Graceffo, A. 'China's National Champions: State Support Make Chinese Companies Dominant' (2017) Foreign Policy Journal https://www.foreignpolicyjournal.com/2017/05/15/chinas-national-champions-state-support-makes-chinese-companies-dominant/

Gunning, D. Stefik, M. Choi, J. Miller, T. Stumpf, S. Yang, G-Z. 'XAI – Explainable Artificial Intelligence' (2019) Science Robotics 4(37)

Gupta, A. 'The Evolution of Fraud: Ethical Implications in the Age of Large-scale Data Breaches and Widespread Artificial Intelligence Solutions Deployment' (2018) International Telecommunications Union Journal 1 <https://www.researchgate.net/profile/Abhishek_Gupta193/publication/323857997_The_Evolution_of_Fraud_Ethical_Implications_in_the_Age_of_Large-Scale_Data_Breaches_and_Widespread_Artificial_Intelligence_Solutions_Deployment/links/5aaffd3f0f7e9b4897c1d066/The-Evolution-of-Fraud-Ethical-Implications-in-the-Age-of-Large-Scale-Data-Breaches-and-Widespread-Artificial-Intelligence-Solutions-Deployment.pdf>

Hacker, P. (2018) Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law. Common Market Law Review. 55, 1143–1186

Hannah-Moffatt, K. 'Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates' (2018) Theoretical Criminology <https://journals.sagepub.com/doi/pdf/10.1177/1362480618763582>

Hauer, T. 'Society and the Second Age of Machines: Algorithms vs Ethics' (2018) Society 55(2) 100-106 100

Heale, R. Twycross, A. 'What is a case study' (2018) Evidence Based Nursing 21(1) 7-8

Hon. Justice Michael Kirby, 'The fundamental problem of regulating technology' (2009) The Indian Journal of Law and Technology, 5 1-32

Hutchinson, T. 'The Doctrinal Method: Incorporating Interdisciplinary Methods in Reforming the Law' (2015) Erasmus Law Review 3 130-138

Huchinson, T. Duncan, N. 'Defining and Describing What We Do: Doctrinal Legal Research' (2012) Deakin Law Review 17(1) 83-120

Huq, A. Z. 'Racial Equality in Algorithmic Criminal Justice' (2019) Duke Law Journal 68(6) 1043-1134

Isaak, J. Hanna, M. J. 'User Data Privacy: Facebook, Cambridge Analytica, and Private Protection' (2018) IEEE Computer 51(8) 56-59

Israni, E. 'Algorithmic Due Process: Mistaken Accountability and Attribution in State v Loomis' (2017) Harvard Journal of Law and Technology – JOLT Digest <https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1>

Jackson, B. W. 'Artificial Intelligence and the Fog of Innovation: A Deep-Dive on Governance and the Liability of Autonomous Systems' (2019) Santa Clara High Technology Law Journal 35(4) 35-63 54

Jackson, M. C. 'Artificial Intelligence and Algorithmic Bias: The Issues With Technology Reflecting History & Humans' (2021) Journal of Business & Technology Law 16(2) 299-316, 309

Janssen, M. van der Voort, H. Wahyudi, A. (2017) Factors Influencing Big Data Decision-Making Quality. Journal of Business Research 70, 338–345. Available at: http://www.sciencedirect.com/science/article/pii/S0148296316304945)

Jiang, F. Jiang, Y. Zhi, H. Dong, Y. Li, H. Ma, S. Wang, Y. Dong, Q. Shen, H. Wang, Y. 'Artificial intelligence in healthcare: past, present and future' (2017) Stroke and Vascular Neurology <https://svn.bmj.com/content/svnbmj/2/4/230.full.pdf>

Johnson, J. 'Artificial intelligence & future warfare: implications for international security' (2019) Defense & Security Analysis 35(2) 147-169

Joseph, J. Turksen, U. 'Harnessing AI for due diligence in CBI Programmes' (2022) Journal of Ethics and Legal Technologies 4(2)

Katyal, S. K. 'Private Accountability in the Age of Artificial Intelligence' (2019) UCLA Law Review 66(1) 54-142

Kazim, E. Almeida, D. Kingsman, N. Kerrigan, C. Koshiyama, A. Lomas, E. Hilliard, A. 'Innovation and opportunity: review of the UK's national AI strategy' (2021) Discover Artificial Intelligence 1(14)

Lambrecht, A. Tucker, C. 'Gender-Based Discrimination in the Display of STEM Career Ads' (2019) Management Science 65(7) <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2018.3093>

Larsson, S. Heintz, F. 'Transparency in artificial intelligence' (2020) Internet Policy Review, 9(2) referencing a study included within Jobin, A. Ienca, M. Vayena, E. 'The global landscape of AI ethics' (2019) Nature Machine Intelligence, 1 389-399

Lepri, B. Oliver, N. Letouze, E. Pentland, A. Vinck, P. 'Fair, Transparent and Accountable Algorithmic Decision-making Processes' (2018) Philosophy & Technology 31, 611-627

Lewis, D. 'International Legal Regulation of the Employment of Artificial-Intelligence-Related Technologies in Armed Conflict' (2020) Moscow Journal of International Law 2 53-64

Lintvedt, M. N. 'Putting a price on data protection infringement' (2021) International Data Privacy Law 12(1)

Lutz,C. 'Digital inequalities in the age of artificial intelligence and big data' (2019) Human Behaviour and Emerging Technology, 1 141-148, 144

MacSíthigh, D. Siems, M. 'The Chinese Social Credit System: A model for other countries?' (2019) Modern Law Review 82(6)

Maedche, A. Legner, C. Benlian, A. Berger, B. Gimpel, H. Hess, T. Hinz, O. Morana, S. Söllner, M. (2019) AI-based Digital Assistants: Opportunities, Threats, and Research Perspectives. Business and Information Systems Engineering. 61(4), 535–544

Mahler, T. 'Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal' (2022) Nordic Yearbook of Law and Informatics < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4001444>

Mann, M. Matzner, M. (2019) Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination. Big Data & Society. Available at: https://journals.sagepub.com/doi/10.1177/2053951719895805

Matthew, C. Halliday, D. 'Big Data and the Liberal Conception of Education' (2017) Theory and Research in Education 15(3) 290-306
Morgan, S. 'Fake news, disinformation, manipulation, and online tactics to undermine democracy' (2018) Journal of Cyber Policy 3(1) 39-43

Orucu, E. 'Methodological Aspects of Comparative Law (2006) European Journal of Law Reform 8(1) 29-42

Olhede, S. C. Wolfe, P. J. 'The growing ubiquity of algorithms in society: implications, impacts and innovations' (2018) Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376(2128) <https://doi.org/10.1098/rsta.2017.0364>

Osoba, O. A. Wesler IV, W. 'The Risks of Artificial Intelligence to Security and the Future of Work' (2017) Perspective: Expert insights on a timely policy issue <https://pdfs.semanticscholar.org/b775/bd1807572e45636e8508667edb5c9fd8cc72.pdf>

Parnas, D. 'The real risks of artificial intelligence' (2017) Communications of the ACM 60(10) 27-31 27

Piccolo, R. ‘AI in criminal sentencing: A risk to our human rights?’ (2018) Bulletin (Law Society of South Australia) 40(11) 15-17

Pathak, V. Jena, B. Kalra, S. ‘Qualitative research’ (2013) Perspectives in Clinical Research 4(3) 192

Racine, E. Boehlen, W. Sample, M. ‘Healthcare uses of artificial intelligence: Challenges and opportunities for growth’ (2019) Healthcare Management Forum 32(5)

Reddy, S. Allan, S. Coghlan, S. Cooper, P. ‘A governance model for the application of AI in health care’ (2019) Journal of the American Medical Informatics Association 27(3) 491-497

Reddy, S. Fox, J. Purohit, M. P. ‘Artificial Intelligence-enabled healthcare delivery’ (2018) Journal of the Royal Society of Medicine 112(1) 22-28 22

Rehman, O.U, Ryan, M. J. ‘On the Dynamics of Design of Future-Proof Systems’ 25th Annual INCOSE International Symposium (2015) DOI: 10.1002/j.2334-5837.2015.00050.x.

Richardson, J. P. Smith, C. Curtis, S. Watson, S. Zhu, X. Barry, B. Sharp, R. R. ‘Patient apprehensions about the use of artificial intelligence in healthcare’ (2021) NPJ Digital Medicine 4 https://doi.org/10.1038/s41746-021-00509-1

Roberts, H. Cowls, J. Morley, J. Taddeo, M. Wang, V. Floridi, L. ‘The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation’ (2021) AI & Society, 36, 59-77

Rolnik, D. et al, ‘Tackling Climate Change with Machine Learning’ (2019) Computers and Society https://arxiv.org/abs/1906.05433

Salehijam, M. ‘The Value of Sysetmic Content Analysis in Legal Research’ (2018) Tilburg International Law Review 23(1-2) 34

Salemink, K. Strijker, D. Bosworth, G. ‘Rural development in the digital age: A systematic literature review on unequal ICT availability, adoption, and use in rural areas’ (2017) Journal of Rural Studies 54 360-371

Samek, W. Muller, K-R. ‘Towards Explainable Artificial Intelligence’ in W. Samek, G. Montavon, A. Vedaldi, L. Hansen, K-R Muller (eds) Explainable AI: Interpreting, Explaining and Visualising Deep Learning (Springer, 2019) 5-22

Savage, N. ‘The race to the top among the world’s leaders in artificial intelligence’ (2020) Nature Index <https://www.nature.com/articles/d41586-020-03409-8>

Scherer, M. U. ‘Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies’ (2016) Harvard Journal of Law and Technology 29(2) 354-400
Shrestha, Y.R. Yang, Y. (2019) Fairness in Algorithmic Decision-Making: Applications in Multi-Winner Voting Machine Learning, and Recommender Systems. Algorithms. 12(9), 199–227.

Sullivan, C. ‘GDPR Regulation of AI and Deep Learning in the Context of IoT Data Processing – A Risky Strategy’ (2018) Journal of Internet Law 22(6) 1-18

Taylor, E. ‘Autonomous Vehicle Decision-Making Algorithms and Data-Driven Mobilities in Networked Transport Systems’, Contemporary Readings in Law and Social Justice 13(1) 9

Tene, O. Polonetsky, J. (2013) Big Data for All: Privacy and User Control in the Age of Analytics. Northwestern Journal of Technology and Intellectual Property. 11(5), 239–273

Theodorou, A. Wortham, R. H. Bryson, J. J. 'Designing and implementing transparency for real time inspection of autonomous robots' (2016) Connection Science 29(3) 230-241

Tolan, S. (2019) Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges. Available at: https://arxiv.org/abs/1901.04730

Veale, M. Borgesius, F. Z. 'Demystifying the Draft EU Artificial Intelligence Act' (2021) Computer Law Review International, 4 97-112

Vick, D. 'Interdisciplinarity and the Discipline of Law' (2004) Journal of Law and Society 31(2) 163-193

Vijayasri, G. V. 'The Importance of International Trade in the World' (2013) International Journal of Marketing, Financial Services & Management Research 2(9) 111-119

Wachter, S. Mittelstadt, B. (2019) A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. Columbia Business Law Review. 2019(2), 494–620

Welch, R. 'Defining Artificial Intelligence' (2019) Society of Motion Picture &Television Engineers Motion Imaging Journal 128(1) 26-33

Wellner, G. Rothman, T. (2020) Feminist AI: Can We Expect Our AI Systems to Become Feminist? Philosophy & Technology. 33, 191–205

Wils, W. P. J. 'E.C. Competition Fines: To Deter or Not to Deter' (1995) Yearbook of European Law 15(1) 17

Wilson, R. 'Cambridge Analytica, Facebook, and Influence Operations: A Case Study and Anticipatory Ethical Analysis' (2019) European Conference on Cyber Warfare and Security 587-595
<https://search.proquest.com/docview/2261006731/fulltextPDF/6234C4E0A0D845B3PQ/1?accountid=12118>

Wirth, N. 'Hello marketing, what can artificial intelligence help you with? (2018) International Journal of Market Research 60(5) 435-439

Wolff, J. Atallah, N. 'Early GDPR Penalties: Analysis of Implementation and Fines Through May 2020' (2021) Journal of Information Policy 11 63-103

Zeng, J. 'Artificial Intelligence and China's authoritarian governance' (2020) International Affairs 96(6) 1441-1459 1442
Zerilli, J. Knott, A. Maclauri, J. Gavaghan, C. (2019) Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? Philosophy & Technology. 32, 661–683

Zhu, J. 'AI ethics with Chinese characteristics? Concerns and preferred solutions in Chinese academia' (2022) AI & Society < https://link.springer.com/article/10.1007/s00146-022-01578-w>

Žliobaitė, I. (2015) A Survey on Measuring Indirect Discrimination in Machine Learning. Available at: https://arxiv.org/abs/1511.00148

Blogs and Websites

Ada Lovelace Institute, AI Council, 'Exploring legal mechanisms for data stewardship' (adalovelaceinstitute.org, 2021) < https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/>

Ahuja, A. ‘Tech luminaries’ beliefs need further examination’ (Financial Times, 2023) <https://www.ft.com/content/edc30352-05fb-4fd8-a503-20b50ce014ab >

AI for Good, ‘About’ (aiforgood.itu.int, 2023) < https://aiforgood.itu.int/about-ai-for-good/>

Angwin, J. Larson, J. Mattu, S. Kirchner, L. ‘Machine Bias: There’s Software Used across the Country to Predict Future Criminals. And It’s Biased against Blacks.' (propublica.org, 2016) < https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Anzini, M. ‘EIPA Briefing 2021/5, The Artificial Intelligence Act Proposal and its implications for Member States’ (eipa.eu, 2021) < https://www.eipa.eu/publications/briefing/the-artificial-intelligence-act-proposal-and-its-implications-for-member-states/>

Baldini, D., ‘Article 22 GDPR and prohibition of discrimination. An outdated provision?’ (cyberlaws.it, 2019) <https://www.cyberlaws.it/2019/article-22-gdpr-and-prohibition-of-discrimination-an-outdated-provision>

Berg, T. Burg, V. Gombović, A. ‘On the Rise of the FinTechs — Credit Scoring Using Digital Footprints’ (fcdic.gov, 2018)< https://www.fdic.gov/bank/analytical/cfr/2018/wp2018/cfr-wp2018-04.pdf >

British Standards Institute (BSI), ‘Standards and Legislation’ (bsigroup.com) <https://www.bsigroup.com/en-GB/standards/Information-about-standards/standards-and-regulation/>

Cabinet Office, ‘Guide to making legislation’ (GOV.UK, 2017) <https://www.gov.uk/government/publications/guide-to-making-legislation>

Cambridge Dictionary, ‘End user’ (dictionary.cambridge.org, 2022) <https://dictionary.cambridge.org/dictionary/english/end-user>

Catapult: High Value Manufacturing, ‘Manufacturing the Future Workforce’ (hvm.catapult.org, 2021) <https://hvm.catapult.org.uk/mtfw/>

Cave, S. OhEigeartaigh S., ‘An AI Race for Strategic Advantage: Rhetoric and Risks’ (2018) Association for the Advancement of Artificial Intelligence <http://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_163.pdf>

Center for Security and Emerging Technology, ‘Ethical norms for New Generation Artificial Intelligence Released’ (cset.georgetown.edu, 2021) <https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/>

Center for Security and Emerging Technology, ‘White Paper on Trustworthy Artificial Intelligence’ (cset.georgetown.edu, 2021) < https://cset.georgetown.edu/publication/white-paper-on-trustworthy-artificial-intelligence/>

Centre for Data, Ethics and Innovation, ‘The roadmap to an effective AI assurance ecosystem’ (GOV.uk, 2021) <https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem>

CIFAR, ‘Pan-Canadian National AI Strategy’ (cifar.org) < https://cifar.ca/ai/>

Circiumaru, A. ‘Three proposals to strengthen the EU Artificial Intelligence Act’ (2021) Ada Lovelace Institute <https://www.adalovelaceinstitute.org/blog/three-proposals-strengthen-eu-artificial-intelligence-act/#:~:text=The%20European%20Commission's%20Artificial%20Intelligence,in%20law%20enforcement 2C%20education%20and>

Davide, B, 'Article 22 GDPR and Prohibition of Discrimination. An Outdated Provision?' (CyberLaws.it, 2019) <https://www.cyberlaws.it/2019/article-22-gdpr-and-prohibition-of-discrimination-an-outdated-provision/>

Degallaix, L. Eliantonio, M., European Environmental Citizens Organisation for Standardisation 'The use of standards in legislation and policies' (ecostandard.org, 2018) <https://ecostandard.org/wp-content/uploads/2018-06-11-The-use-of-standards-in-legislation-and-policies-ECOS-discussion-paper.pdf>

Diplo, 'Artificial Intelligence in Africa: Continental policies and initiatives' (diplomacy.edu, 2022) < https://www.diplomacy.edu/resource/report-stronger-digital-voices-from-africa/ai-africa-continental-policies/>

Dobson, S. J. Ciclitira, K. 'A new regulatory regime and a new regulator: a new era for the regulation of construction products in the UK?' (kennedyslaw.com, 2022) <https://kennedyslaw.com/thought-leadership/blogs/product-safety-blog-in-safe-hands/a-new-regulatory-regime-and-a-new-regulator-a-new-era-for-the-regulation-of-construction-products-in-the-uk/>

Eberle, E. J. 'The Methodology of Comparative Law' (2011) Roger Williams University Law Review 16(1) 51-72

Edelman, '2019 Edelman AI Survey' (Edelman.com, 2019) <https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_AI_Survey_Whitepaper.pdf?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosfutureofwork&stream=future>

EDRi, 'The EU AI Act and fundamental rights: Updates on the political process' (edri.org, 2022) < https://edri.org/our-work/the-eu-ai-act-and-fundamental-rights-updates-on-the-political-process/>

Edwards, J. L. 'An Examination of Consumers' Social Media Trust in the Wake of the Facebook and Cambridge Analytica Scandal' (cache.kzoo.edu, 2019) <https://cache.kzoo.edu/handle/10920/36677>

Edwards, L. 'The EU AI Act: a summary of its significance and scope' (adalovelaceinstitute.org, 2022) <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf >

Eggers, W. D. Turley, M. Kamleshkumar Kishnani, P. 'The future of regulation: Principles for regulating emerging technologies' (Deloitte Insights, 2018) <https://www2.deloitte.com/us/en/insights/industry/public-sector/future-of-regulation/regulating-emerging-technology.html>

Egress Software Technologies, 'The Future of AI in Data Protection: What do the Experts Say?' (egress.com, 2018) <https://www.egress.com/artificial-intelligence-for-data-protection>

Electronic Privacy Information Centre, 'Algorithms in the Criminal Justice System: Pre-Trial Risk Assessment Tools' (epic.org, 2019) <https://epic.org/algorithmic-transparency/crim-justice/>

Equality and Human Rights Commission, 'Article 14: Protection from Discrimination' (equalityhumanrights.com, 2018) <https://www.equalityhumanrights.com/en/human-rights-act/article-14-protection-discrimination>

ETSI, 'Membership of ETSI' (etsi.org, 2021) <https://www.etsi.org/membership/members>

ETSI, ‘Consumer IoT Security’ (ETSI.org) https://www.etsi.org/technologies/consumer-iot-security

Exactpro, ‘Exactpro Test and Automation Approaches: A Case Study in Market Surveillance’ (exactpro.com, 2022) <https://exactpro.com/case-study/market-surveillance#:~:text=The%20optimisation%20of%20the%20Market,requirements%20are%20met%20and%20documented.>

Financial Conduct Authority, ‘Transparency’ (fca.org.uk, 2022) <https://www.fca.org.uk/about/transparency>

Free, R. Kerrigan, C. Zapisetskayac, B. ‘AI, Machine Learning & Big Data Laws and Regulations 2022, United Kingdom’ (2022) Global Legal Insights < <https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/united-kingdom#chaptercontent3>

Future of Life Institute, ‘The Asilomar AI Principles’ (futureoflife.org, 2017) < https://futureoflife.org/open-letter/ai-principles/>

Good, L. Buford, E. ‘Modernizing and Investing in Workforce Development’ (Corporation for a Skilled Workforce, 2021) <https://skilledwork.org/wp-content/uploads/2021/03/Modernizing-and-Investing-in-Workforce-Development.pdf>

Hayes, E. Wall, S, ‘The legal risks of automated decision-making’ (peoplemanagement.co.uk, 2020) https://www.peoplemanagement.co.uk/experts/legal/the-legal-risks-of-automated-decision-making

Health and Safety Executive for Northern Ireland, ‘Declaration of Conformity’ (hseni.gov.uk, 2022) <https://www.hseni.gov.uk/articles/declaration-conformity>

Hewson, V. Turnbridge, J. ‘Who regulates the regulators?’ (2020) Institute of Economic Affairs <https://iea.org.uk/wp-content/uploads/2020/07/Who-regulates-the-regulators_.pdf>

Horton, B. Zeng, J. ‘Can China become the AI superpower?’ (chathamhouse.org, 2021) < https://www.chathamhouse.org/2021/01/can-china-become-ai-superpower>

ICO, ‘Big data, artificial intelligence, machine learning and data protection’ (ico.org.uk, 2017) <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

ICO, ‘Our history’ (ico.org.uk, 2022) < https://ico.org.uk/about-the-ico/our-information/history-of-the-ico/our-history/>

ICO, ‘Who we are’ (ico.org.uk, 2022) < https://ico.org.uk/about-the-ico/who-we-are/>

Jillson, E. ‘Aiming for truth, fairness, and equity in your company’s use of AI’ (ftc.gov, 2021) <https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai>

Jones Day, ‘White Paper, Autonomous Vehicles: Legal and Regulatory Developments in the United States’ (jonesday.com, 2021) < https://www.jonesday.com/-/media/files/publications/2021/05/autonomous-vehicles-legal-and-regulatory-developments-in-the-us/files/autonomous-vehicles-legal-and-regulatory-developme/fileattachment/autonomous-vehicles-legal-and-regulatory-developm.pdf>

Jones, E. Safak, C. (2020) Can Algorithms Ever Make the Grade? Available at: https://www.adalovelaceinstitute.org/can-algorithms-ever-make-the-grade/

Kozyrkov, C. ‘What is AI bias?’ (2019) Towards Data Science <https://towardsdatascience.com/what-is-ai-bias-6606a3bcb814>

Krusenvik, L. ‘Using Case Studies as a Scientific Method: Advantages and Disadvantages’ (2016) Halmstad University < https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1054643&dswid=189>

Lawrence, D. ‘Dynamic Alignment and Regulatory Cooperation between the UK and the EU after Brexit’ (2019) Trade Justice Movement https://www.tjm.org.uk/documents/briefings/TJM-Dynamic-Alignment-and-Regulatory-Cooperation-after-Brexit.pdf

Lawrence, S. V. Martin, M. F. ‘Understanding China’s Political System’, Congressional Research Service (2013) < https://sgp.fas.org/crs/row/R41007.pdf>

Lowry, S. Macpherson, G. (1988) A Blot on the Profession. British Medical Journal. Available at: http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2545288&blobtype=pdf

MacCarthy, M. Propp, K. ‘Machines learn that Brussels writes the rules: The EU’s new AI regulation’ (2021) Lawfare < https://www.lawfareblog.com/machines-learn-brussels-writes-rules-eus-new-ai-regulation>

Maguire, J. ‘Top Performing Artificial Intelligence Companies of 2022’ (datamation.com, 2022) https://www.datamation.com/featured/ai-companies/

Mallapaty, S. ‘China bans cash rewards for publishing papers’ (2020) Nature https://www.nature.com/articles/d41586-020-00574-8

McAdams, R. ‘AI in Africa: Key Concerns and Policy Considerations for the Future of the Continent’ (afripoli.org, 2022) < https://afripoli.org/ai-in-africa-key-concerns-and-policy-considerations-for-the-future-of-the-continent>

Meltzer, J. P. Kerry, C. F. ‘Strengthening international cooperation on artificial intelligence’ (Brookings, 17 Feb 2021) <https://www.brookings.edu/research/strengthening-international-cooperation-on-artificial-intelligence/>

Merchant, G. ‘”Soft Law” Governance of Artificial Intelligence’ (2019) UCLA: The Program on Understanding Law, Science and Evidence (PULSE) https://escholarship.org/uc/item/0jq252ks

Mozur, P. ‘One Month, 500,000 Face Scans: How China Is Using AI to Profile a Minority’ (nytimes.com, 2019) < https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>

National Conference of State Legislatures, ‘Autonomous Vehicle State Bill Tracking Database’ (ncsl.org, 2022) < https://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx>

National Conference of State Legislatures, ‘Legislation Related to Artificial Intelligence’ (2022) NCSL.org <https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx#2021>

National Conference of State Legislatures, ‘Legislation Related to Artificial Intelligence’ (ncsl.org, 2022) https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx

National Conference of State Legislatures, ‘State Legislative Policymaking in an Age of Political Polarization’ (ncsl.org) https://www.ncsl.org/Portals/1/HTML_LargeReports/Partisanship_1.htm

National Cyber Security Centre, ‘Vulnerability Disclosure Toolkit’ (ncsc.gov.uk, 2020) https://www.ncsc.gov.uk/information/vulnerability-disclosure-toolkit

National Security Commission on Artificial Intelligence, 'About us' (nscai.gov, 2022) < https://www.nscai.gov/about/>

New York State Department of Financial Services, 'Report on Apple Card Investigation' (2021) <https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf> accessed 20/03/2023

Newman, J. 'Explainability won't save AI' (brookings.edu, 19 May 2021) <https://www.brookings.edu/techstream/explainability-wont-save-ai>

NIST, 'AI Risk Management Framework: Initial Draft' (2022) < https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf>

NIST, 'Technical AI standards' (nist.gov, 6 August 2021) <https://www.nist.gov/artificial-intelligence/technical-ai-standards>

Nkwanyana, K. 'China's AI deployment in Africa poses risks to security sovereignty' (aspistrategist.org, 2021) < https://www.aspistrategist.org.au/chinas-ai-deployment-in-africa-poses-risks-to-security-and-sovereignty/>

OECD, Artificial Intelligence in Society (OECD Publishing, Paris 2019) <https://doi.org/10.1787/eedfee77-en>

OECD, 'Better Regulation in Europe: Ireland' (oecd.org, 2010) <https://www.oecd.org/ireland/betterregulationineuropeireland.htm>

OECD, 'Industry Self-Regulation: Role and Use in Supporting Consumer Interests' (2015) Directorate for Science, Technology and Innovation Committee on Consumer Policy <http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/CP(2014)4/FINAL&docLanguage=En>

OECD, 'Regulatory Policy in the Slovak Republic: Towards Future-Proof Regulation' (2020), OECD Reviews of Regulatory Reform (OECD Publishing, Paris) <https://www.oecd-ilibrary.org/sites/94d061e5-en/index.html?itemId=/content/component/94d061e5-en>

OECD.AI, 'Database of national AI policies (oecd.ai, 2022) https://oecd.ai/en/dashboards/countries/UnitedKingdom

Ofcom, 'What is Ofcom?' (ofcom.org.uk, 2022) < https://www.ofcom.org.uk/about-ofcom/what-is-ofcom>

Oloruntade, G. Omoniyi, F. 'Where is Africa in the global conversation on regulating AI?' (techcabal.com, 2023) < https://techcabal.com/2023/05/26/where-is-africa-in-the-global-conversation-on-regulating-ai/>

Organisation for Economic Co-operation and Development (OECD), 'OECD Council Recommendation on Artificial Intelligence' (2019) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

Privacy Policies, 'Implementing Privacy by Design' (privacypolicies.com, 5 January 2021) <https://www.privacypolicies.com/blog/privacy-by-design/#What_Is_Privacy_By_Design>

Privacy Policies, 'Implementing Privacy by Design' (privacypolicies.com, 5 January 2021) <https://www.privacypolicies.com/blog/privacy-by-design/#What_Is_Privacy_By_Design>

PWC, 'Eight emerging technologies and six convergence themes you need to know about' (2022) < https://www.pwc.com/us/en/tech-effect/emerging-tech/essential-eight-

technologies.html#:~:text=They%20include%3A%20artificial%20intelligence%20(AI,pandemic%20accelerating%20emerging%20tech%20adoption.>

Radwan, S. 'Egypt's Ai strategy is more about development than AI' (oecd.ai, 2021) < https://oecd.ai/fr/wonk/egypt-ai-strategy>

Robin, S. (2017) Opening the Lid on Criminal Sentencing Software. Duke Today. Available at: https://today.duke.edu/2017/07/opening-lid-criminal-sentencing-software

Roetzer, P. 'Funding for AI Sales and Marketing Companies Exceeds $5.2 Billion' (Marketing Artificial Intelligence Institute, 8 Jan 2019) <https://www.marketingaiinstitute.com/blog/funding-for-ai-powered-sales-and-marketing-companies-exceeds-5.2-billion>

Saïd Business School (University of Oxford) 6 week, online Artificial Intelligence programme targeted at understanding AI for business https://oxford-onlineprogrammes.getsmarter.com/presentations/lp/oxford-artificial-intelligence-programme

Savage, N. 'The race to the top among the world's leaders in artificial intelligence' (2020) Nature Index <https://www.nature.com/articles/d41586-020-03409-8>

Shaddick, G. 'COP26 and beyond: the crucial role for AI in tackling climate change' (turing.ac.uk, 2021) <https://www.turing.ac.uk/blog/cop26-and-beyond-crucial-role-ai-tackling-climate-change>

Sharwood, S. 'Russian ChessBot breaks child opponent's finger' (theregister.com, 2022) https://www.theregister.com/2022/07/25/russian_chessbot_breaks_players_finger/

Sheehan, M. 'China's New AI Governance Initiatives Shouldn't Be Ignored' (carnegieendowment.org, 2022) < https://carnegieendowment.org/2022/01/04/china-s-new-ai-governance-initiatives-shouldn-t-be-ignored-pub-86127>

Sheikhansari, I. 'The Impact of AI on Social Class and Jobs: A Closer Look' (linkedin.com, 2023) <https://www.linkedin.com/pulse/impact-ai-social-class-jobs-closer-look-iman-sheikhansari/>

SHERPA, 'Guidelines for the Ethical Use of AI and Big Data Systems', and 'Guidelines for the Ethical Development of AI ad Big Data Systems: An Ethics by Design approach' (project-sherpa.eu, 2020) < https://www.project-sherpa.eu/guidelines/>

Simpson, L. 'Looking at AI through a global, gender lens' (medium.com, 2019) <https://medium.com/hellobrink-co/looking-at-ai-through-a-global-gender-lens-f12aa92c55a4>

Smith, R. A. (2017) Opening the lid on criminal sentencing software. Available at: https://today.duke.edu/2017/07/opening-lid-criminal-sentencing-software

Sonnad, N. (2017) Google Translate's Gender Bias Pairs "He" with "Hardworking" and "She" with Lazy, and Other Examples. Available at: https://qz.com/1141122/google-translates-gender-bias-pairs-he-with-hardworking-and-she-with-lazy-and-other-examples/

Statista, 'Annual net sales revenue of Amazon from 2004-2021' (statista.com, 2022) https://www.statista.com/statistics/266282/annual-net-revenue-of-amazoncom/

Statista, 'Distribution of the gross domestic product (GDP) in China in 2021' (statista.com, 2022) < https://www.statista.com/statistics/1124008/china-composition-of-gdp-by-industry/#:~:text=In%202021%2C%20the%20industrial%20sector,of%20the%20country's%20economic%20output.>

Statista, ‘Largest fines issued for General Data Protection Regulation (GDPR) violations as of July 2022’ (statista.com, 2022) https://www.statista.com/statistics/1133337/largest-fines-issued-gdpr/

Strubell, E. Ganesh, A. McCallum, A. ‘Energy and Policy Considerations for Deep Learning in NLP’ (Cornell University 2019) < https://arxiv.org/abs/1906.02243>

Sussman, H. McKenney, R. Wolfington, A. ‘U.S. Artificial Intelligence Regulation Takes Shape’ (Orrick LLP, 2021) https://www.orrick.com/en/Insights/2021/11/US-Artificial-Intelligence-Regulation-Takes-Shape#:~:text=Artificial%20Intelligence%20(AI)%20has%20the,next%20era%20of%20technological%20advancement.

Squicciarini, M. Nachtigall, H. ‘Demand for AI skills in jobs: Evidence from online job postings’ (2021) OECD Science, Technology and Industry Working Papers < https://www.oecd-ilibrary.org/docserver/3ed32d94-en.pdf?expires=1643041720&id=id&accname=guest&checksum=24A35C03F178E06818B5B093A47BF001>

Tech Republic, ‘Top 10 AI skills and how to get them’ <https://www.techrepublic.com/article/here-are-the-10-most-in-demand-ai-skills-and-how-to-develop-them/>

Tessian, ’30 Biggest GDPR Fines So Far (2020, 2021, 2022)’ (tessian.com, 2022) https://www.tessian.com/blog/biggest-gdpr-fines-2020/

The Alan Turing Institute, ‘New UK initiative to shape global standards for artificial intelligence’ (turing.ac.uk, 2022) https://www.turing.ac.uk/news/new-uk-initiative-shape-global-standards-artificial-intelligence

The Alan Turing Institute, ‘Women in Data Science and AI: Project Aims’ (turing.ac.uk, 2023) <https://www.turing.ac.uk/research/research-projects/women-data-science-and-ai-new#:~:text=Women%20make%20up%20half%20the,online%20and%20physical%20workplace%20cultures.>

The Community Research Development Information Service (CORDIS) (2020) Safeguarding Equality in the European Algorithmic Society: Tackling Discrimination in Algorithmic Profiling through EU Equality Law. Available at: https://cordis.europa.eu/project/id/898937 (Accessed 26 August 2020)

The Community Research Development Information Service (CORDIS) ‘Safeguarding Equality in the European Algorithmic Society: Tackling Discrimination in Algorithmic Profiling through EU Equality Law’ (2020) <https://cordis.europa.eu/project/id/898937>

The Joint Council for the Welfare of Immigrants (2020) We Won! Home Office to Stop Using Racist Visa Algorithm. Available at: https://www.jcwi.org.uk/news/we-won-home-office-to-stop-using-racist-visa-algorithm

The Law Library of Congress, ‘Regulation of Artificial Intelligence in Selected Jurisdictions’ (2019) < https://tile.loc.gov/storage-services/service/ll/llglrd/2019668143/2019668143.pdf>

The Law Library of Congress, ‘Regulation of Artificial Intelligence in Selected Jurisdictions’ (2019) < https://tile.loc.gov/storage-services/service/ll/llglrd/2019668143/2019668143.pdf>

Think Tank, European Parliament, ‘General product safety regulation’ (europarl.europa.eu, 2021) < https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698028>

Trump Whitehouse Archives, ‘Artificial Intelligence for the American People’ < https://trumpwhitehouse.archives.gov/ai/executive-order-ai/>

UKAS, ‘Laboratory Accreditation – Biological’ (UKAS.com, 2022) < https://www.ukas.com/accreditation/standards/laboratory-accreditation/biological/>

UNESCO, ‘Recommendations on the Ethics of Artificial Intelligence’ (unesdoc.unesco.org, 2022) <https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi>

UNESCO, ‘UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence’ (unesco.org, 2021) <https://www.unesco.org/en/articles/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence>

UNI Global Union, ‘Top 10 Principles for Ethical Artificial Intelligence’ (thefutureworldofwork.org, 2017) < http://www.thefutureworldofwork.org/opinions/10-principles-for-ethical-ai/>

UNICRI, ‘Centre on Artificial Intelligence and Robotics’ (unicri.it, 2023) < https://unicri.it/topics/ai_robotics/centre>

Whittaker, M. Alper, M. Bennett, C.L. Hendren, S. Kaziunas, L. Mills, M. Morris, M.R. Rankin, J. Rogers, E. Salas, M. West, S.M. (2019) Disability, Bias, and AI. Available at: https://ainowinstitute.org/disabilitybiasai-2019.pdf?fbclid=IwAR31dX3o_nkVf-cirQ9P-yJqRRkT1vcKU3MgcEAeWVwUgA0Ue1c-60Zd9OE

Media Coverage

Angelova, M. McCluskey, M. ‘Chess-playing robot breaks boy’s finger at Moscow tournament’ (edition.cnn.com, 2022) <https://edition.cnn.com/2022/07/25/europe/chess-robot-russia-boy-finger-intl-scli/index.html>

Baidu, ‘App cannot force personalised recommendation, do you know how to close it?’ (baijiahao.baidu.com, 2021) <https://baijiahao.baidu.com/s?id=1709066811285781912&wfr=spider&for=pc>

BBC, Apple’s “Sexist” Credit Card Investigated by US Regulator (BBC.co.uk, 2019) <https://www.bbc.co.uk/news/business-50365609 >

Cadwalladr, C. ‘The Cambridge Analytica Files’ The Guardian (London, 18th March 2018) <http://davelevy.info/Downloads/cabridgeananalyticafiles%20-theguardian_20180318.pdf>

Conrad, J. Knight, W. ‘China is About to Regulate AI – and the World is Watching’ (wired.com, 2022) < https://www.wired.com/story/china-regulate-ai-world-watching/>

Dewey, C. ’98 personal data points that Facebook uses to target ads at you’ The Washington Post (washingtonpost.com, 2016) <https://www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you/>

Goodier, M., ‘Top A-Level Grades Soar at Private Schools as Sixth Form Colleges Lose Out’ (newstatesman.com, 2020) https://www.newstatesman.com/politics/education/2020/08/top-level-grades-soar-private-schools-sixth-form-colleges-lose-out

Harris, J. Ammanath, B. ‘Defining trustworthy AI’ (2022) Towards Data Science Podcast Transcript <https://towardsdatascience.com/defining-trustworthy-ai-234a97c39035>

Harwell, D. ‘Both Democrats and Republicans blast facial-recognition technology in a rare bipartisan moment’ (washintonpost.com, 2019) <https://www.washingtonpost.com/technology/2019/05/22/blasting-facial-recognition-technology-lawmakers-urge-regulation-before-it-gets-out-control/>

Knight, W. (2019) The Apple Card Didn’t ‘See’ Gender—And That’s the Problem Available at: https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/

Noujaim, J. Amer, K. The Great Hack (Netflix 2019)

Ormond, E. 'Artificial intelligence In South Africa comes with special dilemmas – plus the usual risks' (2023) The Conversation <https://theconversation.com/artificial-intelligence-in-south-africa-comes-with-special-dilemmas-plus-the-usual-risks-194277>

Oxford Insights, 'Government Artificial Intelligence Readiness Index' (2019) < https://africa.ai4d.ai/wp-content/uploads/2019/05/ai-gov-readiness-report_v08.pdf>

Pooya, G. 'The great AI debate: What candidates are (finally) saying about artificial intelligence' (thehill.com, 2019) < https://thehill.com/opinion/technology/473794-the-great-ai-debate-what-candidates-are-finally-saying-about-artificial/>

The Guardian (2020) A-Level Results Day 2020 Live (theguardian.com, 2020) https://www.theguardian.com/education/live/2020/aug/13/a-level-results-day-2020-live-students-teachers-government-ucas-mock-exams-triple-lock-nick-gibb#:~:text=A-levels-, A-level%20results%20day%202020%20live%3A%2039.1%25%20of%20pupils, England%20downgraded%20-%20as%20it%20happened&text=That's%20all%20from%20me%2C%20Caroline%20Davies.&text=More%20than%20a%20third%20of, were%20downgraded%20by%20three%20grades

The New York Times, Apple Card Investigated after Gender Discrimination Complaints (nytimes.com, 2019) https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html

The Telegraph, A-Level and GCSE Results Update: How Are 2020 Grades Being Calculated without Exams? (telegraph.co.uk, 2020) Available at: https://www.telegraph.co.uk/education-and-careers/2020/08/18/a-level-gcse-results-grades/

Vincent, J. 'Robots and AI are going to make social inequality even worse, says new report' (2017) < https://www.theverge.com/2017/7/13/15963710/robots-ai-inequality-social-mobility-study>

Wellsteed-Crook, E. 'Regulate tech to realise the benefits' NewStatesman (newstatesman.com, 2020) <https://www.newstatesman.com/spotlight/emerging-technologies/2020/09/regulate-tech-realise-benefits>

Reports and Government Publications

African Commission on Human and Peoples' Rights, 'Declaration of Principles on Freedom of Expression and Access to Information in Africa' (achpr.au.int, 2019) < https://achpr.au.int/en/node/902#:~:text=The%20Declaration%20establishes%20or%20affirms,to%20express%20and%20disseminate%20information.>

African Union Development Agency, 'The African Union Artificial Intelligence Continental Strategy for Africa' (nepad.org, 2022) < https://www.nepad.org/news/african-union-artificial-intelligence-continental-strategy-africa>

African Union, 'Agenda 2063: The Africa We Want' (au.int, 2023) <https://au.int/en/agenda2063/overview>

African Union, 'The Digital Transformation Strategy for Africa (2020-2030)' (au.int, 2023) < https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf>

Article 29 Data Protection Working Party (2017) Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/670 Adopted on 3

October 2017 As Last Revised and Adopted on 6 February 2018. Available at: http://ec.europa.eu/justice/data-protection/index_en.htm

Chief Executives Board for Coordination, High-Level Committee on Programmes, Inter-Agency Working Group on Artificial Intelligence, 'Principles for the Ethical Use of Artificial Intelligence in the United Nations System' (unsceb.org, 2022) < https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf

Cooperative Governance, Republic of South Africa 'A South African Smart Cities Framework' (cogta.gov.za, 2021) < https://www.cogta.gov.za/cgta_2016/wp-content/uploads/2023/01/Annexure-A-DCoG_Smart-Cities-Framework.pdf>

Cyberspace Administration of China, 'Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms' (2021) No. 7 <https://digichina.stanford.edu/work/translation-guiding-opinions-on-strengthening-overall-governance-of-internet-information-service-algorithms/>

Cyberspace Administration of China, 'Internet Information Service Algorithmic Recommendation Management Provisions (2022) < https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/>

Department for Business, Energy & Industrial Strategy, 'Advanced Research and Invention Agency (ARIA): policy statement' (GOV.uk 2021) <https://www.gov.uk/government/publications/advanced-research-and-invention-agency-aria-statement-of-policy-intent/advanced-research-and-invention-agency-aria-policy-statement>

Department for Business, Energy and Industrial Strategy, 'The Better Regulation Framework' (2020) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/916918/better-regulation-guidance.pdf>

Department for Business, Energy, & Industrial Strategy, Department for Digital, Culture, Media and Sport, Office for Artificial Intelligence, 'Turing Artificial Intelligence Fellowships' (GOV.uk 2021) <https://www.gov.uk/government/publications/turing-artificial-intelligence-fellowships/turing-artificial-intelligence-fellowships>

Department for Digital Culture, Media and Sport, 'Our 10 Tech Priorities' (DCMS.gov.uk) <https://dcms.shorthandstories.com/Our-Ten-Tech-Priorities/index.html>

Department for Digital, Culture, Media and Sport, 'Data: a new direction' (GOV.uk, 2021) <https://www.gov.uk/government/consultations/data-a-new-direction>

Department for Digital, Culture, Media and Sport, 'Government response to the call for views on consumer connected product cyber security legislation' (GOV.uk, 2021) <https://www.gov.uk/government/publications/regulating-consumer-smart-product-cyber-security-government-response/government-response-to-the-call-for-views-on-consumer-connected-product-cyber-security-legislation>

Department for Digital, Culture, Media and Sport, 'Government response to the call for views on consumer connected product cyber security legislation' (GOV.UK, 21 April 2021) <https://www.gov.uk/government/publications/regulating-consumer-smart-product-cyber-security-government-response/government-response-to-the-call-for-views-on-consumer-connected-product-cyber-security-legislation>

Department for Digital, Culture, Media and Sport, 'Initial National Cyber Security Skills Strategy: increasing the UK's cyber security capability – a call for views' (GOV.UK, May 3 2019) <https://www.gov.uk/government/publications/cyber-security-skills-strategy/initial-national-cyber-security-skills-strategy-increasing-the-uks-cyber-security-capability-a-call-for-views-executive-summary>

Department for Digital, Culture, Media and Sport, 'National Data Strategy' (GOV.uk 2020)

Department for Digital, Culture, Media and Sport, 'National Data Strategy Mission 1 Policy Framework: Unlocking the value of data across the economy' (GOV.uk 2021) <https://www.gov.uk/government/publications/national-data-strategy-mission-1-policy-framework-unlocking-the-value-of-data-across-the-economy/national-data-strategy-mission-1-policy-framework-unlocking-the-value-of-data-across-the-economy>

Department for Digital, Culture, Media and Sport, 'New UK initiative to shape global standards for artificial intelligence' (GOV.uk, 2022) <https://www.gov.uk/government/news/new-uk-initiative-to-shape-global-standards-for-artificial-intelligence>

Department for Digital, Culture, Media and Sport, 'World-first online safety laws introduced in Parliament' (GOV.UK, 2022) <https://www.gov.uk/government/news/world-first-online-safety-laws-introduced-in-parliament>

Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, Office for Artificial Intelligence, 'New strategy to unleash the transformational power of Artificial Intelligence' (GOV.UK, 2021) <https://www.gov.uk/government/news/new-strategy-to-unleash-the-transformational-power-of-artificial-intelligence>

Department for Digital, Culture, Media and Sport, Department for Business, Energy, and Industrial Strategy, Office for Artificial Intelligence, 'AI Roadmap' (GOV.UK, 2021) <https://www.gov.uk/government/publications/ai-roadmap/executive-summary>

Department for Digital, Culture, Media and Sport, Office for Artificial Intelligence, 'New UK initiative to shape global standards for Artificial Intelligence' (GOV.uk, 2022)

Department for Education, 'National Skills Fund' (GOV.uk 2021) <https://www.gov.uk/guidance/national-skills-fund#skills-bootcamps>

Department for Education, 'Skills for jobs: lifelong learning for opportunity and growth' (2021) (GOV.uk, 2021) < https://www.gov.uk/government/publications/skills-for-jobs-lifelong-learning-for-opportunity-and-growth>

Department for Science, Innovation and Technology, Office for Artificial Intelligence, 'A pro-innovation approach to AI regulation' (gov.uk, 2023) < https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>

Department of Telecommunications and Postal Services, 'Presidential Commission on Fourth Industrial Revolution' (2019) < https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26873>

Department for Transport, National Highway Traffic Safety Commission, 49 CFR Part 571, Docket No. NHTSA-2021-0003, RIN 2127-AM06 Occupant Protection for Vehicles with Automated Driving Systems

European Commission High Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (ec.europa.eu, 2019) < https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

European Commission, ‘A European approach to artificial intelligence’ (digital-strategy.ec.europa.eu 2022)<https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>

European Commission, ‘EU-Africa: Global Gateway Investment Package’ (commission.europa.eu, 2022) < https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/stronger-europe-world/global-gateway/eu-africa-global-gateway-investment-package_en#accelerating-the-digital-transition>

European Commission, ‘European Commission Staff Working Document: Liability for emerging digital technologies’ (2018) <https://ec.europa.eu/digital-single-market/en/news/european-commission-staff-working-document-liability-emerging-digital-technologies>

European Commission, ‘Fines for breaking EU competition law’ (ec.europa.eu) <https://ec.europa.eu/competition/cartels/overview/factsheet_fines_en.pdf>

European Commission, ‘High-level expert group on artificial intelligence’ (digital-strategy.ec.europa.eu, 23 June 2021) https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

European Commission, ‘Member States and Commission to work together to boost artificial intelligence “made in Europe”’ (digital-strategy.ec.europa.eu, 7 December 2018)

European Commission, ‘Regulatory framework proposal on artificial intelligence’ (digital-strategy.ec.europa.eu, 2022) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

European Commission, ‘Regulatory framework proposal on Artificial Intelligence’ (European Commission, 22 April 2021) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

European Commission, ‘Regulatory framework proposal on artificial intelligence’ (digital-strategy.ec.europa.eu, 2022) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

European Commission, ‘The Cybersecurity Strategy’ (Digital-Strategy.ec.europa.eu, 2022) https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-strategy

European Commission, ‘The Digital Services Act package’ (Digital-Strategy.ac.europa.eu, 2022) < https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package>

European Commission, ‘White Paper on Artificial Intelligence – A European approach to excellence and trust’ COM (2020) 65 final

European Commission, Artificial Intelligence – A European approach to excellence and trust (White Paper (COM 2020) 65 final)

European Data Protection Supervisor, ‘The History of General Data Protection Regulation’ (edps.europa.eu, 2018) <https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en >

European Parliament Committee on Legal Affairs, ‘Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) 31.5.2016

European Union Agency for Fundamental Rights ‘#BigData: Discrimination in Data-Supported Decision Making’ (fra.europa.eu, 2018) https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf

Executive Office of the President, National Science and Technology Council, Committee on Technology, ‘Preparing for the future of artificial intelligence’ (2016)

House of Lords Liaison Committee, ‘AI in the UK: No Room for Complacency’ HL 196 2019-21

House of Lords Library (2020) Predictive and Decision-Making Algorithms in Public Policy. Available at: https://lordslibrary.parliament.uk/research-briefings/lln-2020-0045/

House of Lords Select Committee on Artificial Intelligence, ‘AI in the UK: ready, willing and able?’ (2018) HL Paper 100

International Trade Administration, ‘South Africa – Country Commercial Guide’ (trade.gov, 2023) < https://www.trade.gov/knowledge-product/south-africa-information-technology >

Ipsos Mori, ‘Understanding the UK AI labour market: 2020’ (GOV.uk, 2020) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/984671/DCMS_and_Ipsos_MORI_Understanding_the_AI_Labour_Market_2020_Full_Report.pdf>

Knowledge Exchange Unit, UK Parliament, ‘Research impact on policy’ (GOV.uk 2021) < https://www.parliament.uk/globalassets/assets/teams/post/research_impact_on_policy_briefing_document_june21.pdf?__cf_chl_managed_tk__=ytKbXGUsLt8lItZ6DbgMLGWLVIlYFzGiukHRoCjpuUQ-1643039911-0-gaNycGzNCeU>

Ministry of Communications and Information Technology, ‘Boost your Business - Creativa Innovation Hubs’ (mcit.gov.eg, 2023) < https://mcit.gov.eg/en/Innovation/Boost_your_Business/Creativa_Innovation_Hubs>

National Council for Artificial Intelligence, ‘Egyptian Charter for Responsible AI’ (mcit.gov.eg, 2023) <https://mcit.gov.eg/en/Media_Center/Press_Room/Press_Releases/66939#:~:text=Egypt%20was%20also%20the%20first,Accountability%2C%20and%20Security%20and%20Safety.>

National Council of Artificial Intelligence, ‘Egypt National Artificial Intelligence Strategy’ (2022) < https://mcit.gov.eg/Upcont/Documents/Publications_672021000_Egypt-National-AI-Strategy-English.pdf>

Office for AI, Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, ‘National AI Strategy’ (2021) <https://www.gov.uk/government/publications/national-ai-strategy>

Office for AI, Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, ‘National AI Strategy’ (2021) <https://www.gov.uk/government/publications/national-ai-strategy>

Office for AI, Department for Digital, Culture, Media and Sport, Department for Business, Energy and Industrial Strategy, ‘National AI Strategy’ (2021) <https://www.gov.uk/government/publications/national-ai-strategy>

Office for Product Safety and Standards, ‘ Guidance: Product safety advice for businesses’ (gov.uk, 2021) https://www.gov.uk/guidance/product-safety-advice-for-businesses

Office for Product Safety and Standards, Department for Business, Energy & Industrial Strategy, ‘Designated standards – guidance’ (GOV.UK, 6 January 2021) <https://www.gov.uk/guidance/designated-standards>

Office of Science and Technology Policy, ‘The White House Launches the National Artificial Intelliegnce Initiative Office’ (trumpwhitehouse.archives.gov, 2021) < https://trumpwhitehouse.archives.gov/briefings-statements/white-house-launches-national-artificial-intelligence-initiative-office/>

Office of Science and Technology Policy, 'The White House Launches the National Artificial Intelliegnce Initiative Office' (trumpwhitehouse.archives.gov, 2021) < https://trumpwhitehouse.archives.gov/briefings-statements/white-house-launches-national-artificial-intelligence-initiative-office/>

The Committee on Standards in Public Life, 'Artificial Intelligence and Public Standards' (2020)

U.S. Food and Drug Administration, 'Artificial Intelligence and Machine Learning in Software as a Medical Device' (fda.gov., 2021) https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

U.S. Good and Drug Administration, 'Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML) Based Software as Medical Devices (SaMD), Discussion Paper and Request for Feedback' (fda.gov, 2019) https://www.fda.gov/media/122535/download accessed

US Department of Commerce, 'Department of Commerce Establishes National Artificial Intelligence Advisory Committee' (2021) < https://www.commerce.gov/news/press-releases/2021/09/department-commerce-establishes-national-artificial-intelligence>

Working Group on Artificial Intelligence, 'Mauritius Artificial Intelligence Strategy' (ncb.govmu.org, 2018) <https://ncb.govmu.org/ncb/strategicplans/MauritiusAIStrategy2018.pdf>

Industry Standards and Supplementary Publications

British Standards Institute, 'BS 8611 Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems'(standardsdevelopment.bsigroup.com, 2016) < https://standardsdevelopment.bsigroup.com/projects/2022-00279#/section>

European Telecommunications Standards Institute (ETSI), European Standard (EN) 303 645 v.2.1.1 'Cyber Security for Consumer Internet of Things: Baseline Requirements'

IEEE, 'The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems' (standards.ieee.org, 2017) < https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_general_principles_v2.pdf>

International Telecommunication Union (ITU), 'United nations Activities on Artificial Intelligence (AI)' (aiforgood.itu.int, 20220) < https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/>