

CENTRE INTERNATIONAL DE FORMATION EUROPEENNE

SCHOOL OF GOVERNMENT

INSTITUT EUROPEEN • EUROPEAN INSTITUTE



**Ethical governance for the use of data and AI focusing on the UK and use
in public service delivery**

BY
Sophie Natasha Medd

**A thesis submitted for the Joint Master degree in
Global Economic Governance & Public Affairs (GEGPA)**

Academic year
2020 – 2021

July 2021

Supervisor: Benoit Abeloos
Reviewer: Lorenzo Pupillo

Plagiarism Statement

I hereby declare that I have composed the present thesis autonomously and without use of any other than the cited sources or means. I have indicated parts that were taken out of published or unpublished work correctly and in a verifiable manner through a quotation. I further assure that I have not presented this thesis to any other institute or university for evaluation and that it has not been published before.

30 July 2021 Sophie Natasha Medd

Abstract

The use of data and artificial intelligence (AI) presents many benefits but also risks for society. Many of the risks associated with the use of data and artificial intelligence are moral problems. Ethical issues such as privacy, transparency, accountability, and bias. This study focuses on the UK in an attempt to assess how the current regulatory, legal and governance frameworks are evolving and adapting to remove ethical barriers and protect individuals and groups in relation to data and AI implementation in public services.

The study reviews and breaks down the key ethical problems as they relate to UK public services. An assessment of the current ability and robustness of the existing regulatory and governance frameworks is provided. The study then investigates the evolution of the social innovation process in the nonprofit sector and tests whether this evolving governance structure can provide a solution for the public sector in relation to decision bias; the question of fairness and the need to avoid the negative impact of social rejection in relation to data and AI use in public services.

Results demonstrate the gaps which exist in governance and regulation in relation to ethical barriers and how the ongoing need to access better data (both qualitative and quantitative) could be addressed through a new form of partnership working between the UK public sector and nonprofits. Such a partnership would enable the public sector to leverage the evolution of the Theory of Change within nonprofits, an evidence based, futures and outcomes focused process designed to capture the change journey of social innovation. Specifically as the ToC operates with the lens of socially preferable, as opposed to socially acceptable, in line with the public service equality duty.

Acknowledgements

Throughout the writing of this research, I have received a great deal of support and assistance.

Supervisor Benoit Abeloos and reviewer Lorenzo Pupillo.

I would like to thank all the administration at CIFE and Luiss School of Government for enabling me to complete my Master's programme in the middle of a pandemic.

Thanks to my family for continuing to support me and enabling me to re-enter higher education. Thank you also to my peers, colleagues, and friends from across the public and nonprofit sectors. You provided me with invaluable knowledge and wisdom. Specifically, I would like to say thank you to Samantha Larner, Mark Hazelby, and Andy Melia for the hours spent with me, imparting their expertise and knowledge.

List of Figures	6
List of Abbreviations	7
Introduction to the Research Questions	8
Limitations of the Research	10
Chapter 1	11
Methodology and Approach	11
1.1 Unstructured Interviews	11
1.2 A History of Development	11
1.3 Ethics and AI Debate Areas	12
1.4 UK Case Study	12
1.5 Addressing the Gaps	12
List of Key Definitions	14
Chapter 2	17
The History of AI - Multidisciplinary Approach Combined with Culture	17
2.1 Rage Against the Machine	18
Chapter 3	21
Literature Review - The Key Debate Areas	21
3.1 Privacy and Surveillance	21
3.2 Changing and Directing Behaviours	23
3.3 Transparent Decision Making	25
3.4 Decision Bias	26
3.5 Employment and Automation	29
3.6 Autonomous Systems	32
3.7 Superintelligence brings The Singularity	34
3.8 Summary	35
Chapter 4	37
Ethics, Data and AI - The UK Case Study	37
4.1 Why the UK?	37
4.2 The Nolan Principles - Towards Values-Based Governance	38
4.3 The Current Ecosystem - Key Institutions in the UK	40
4.4 Ethical Guidelines for Data and AI Use in the UK	41
4.5 Data, AI, Ethics and the UK Rule of Law	45

4.6 Review into Bias in Algorithmic Decision Making	46
4.7 Summary	50
Chapter 5	52
Solutions to Algorithmic Bias - The Nonprofits	52
5.1 The rise of the Theory of Change	53
5.2 Key Building Blocks of the Theory of Change	54
5.3 Theory of Change as a Solution to Algorithmic Bias	56
5.4 Summary	58
Chapter 6	59
Conclusion	59
Bibliography	63
Annexes	71
Annex 1- Infographic for a history of AI	71
Annex 2 - Example of a nonprofit Theory of Change (ToC)	73
Annex 3 - Equality Impact Assessment (EIA)	76

List of Figures

Figure	Page
Figure 1 A Rioting Mob of Luddites	19
Figure 2 The SUM Values	42
Figure 3 The FAST Track Principles	43
Figure 4 Process-Based Governance Framework	44

List of Abbreviations

AI - Artificial Intelligence

BEIS - Department for Business, Energy & Industrial Strategy

CDEI - The Centre for Data Ethics and Innovation

DCMS - The Department for Digital, Culture, Media & Sport

EHRC - Equality and Human Rights Commission

GDPR - General Data Protection Regulation

GDS - Government Digital Service

MOT - Ministry of Transport Test

NCVO - National Council for Voluntary Organisations

ICO - The Information Commissioner's Office

PBG - Process-based governance

PSED - Public Sector Equality Duty

ToC - Theory of Change

Introduction to the Research Questions

In the summer of 2020, what is now known simply as "the exam fiasco" took place in the UK

Students were unable to take their A-Level and GCSE exams due to the pandemic. The resolution was the use of an algorithm to determine student grades. The result was accusations of unfair results. Students from disadvantaged backgrounds saw their grades go down, and those in public, fee-paying schools, saw their grades increase. Protests and outcry across stakeholder groups, including students, universities, parents, schools, think tanks, and statisticians, resulted in a Government u-turn. The algorithm was dropped with teacher predictions made before the pandemic used for the grades (Kolkman, 2020).

It is not the first time the UK Government has made a u-turn concerning the use of algorithms. That very same summer of 2020, the UK Home Office was required to drop a system it was using to approve visa applications due to claims of it favouring white applicants, as stated by charity Foxglove.

These two high-profile cases of algorithm use in public services, combined with persistent problems and failures with the technical rollout of the UK Governments' Covid19 Track and Trace system, have created an atmosphere of distrust in the government's technical capabilities. Trust in UK institutions is declining; the 2020 Edelman Trust Barometer has the UK at its lowest position, one place from the bottom spot (Williams, 2020). While not solely due to digital transformation, high-profile scandals that centre around public sector fairness certainly do not aid the return of civic trust.

The Guardian newspaper revealed that after the A-level fiasco, around 20 UK local authorities had consequently dropped the use of an algorithm used to predict welfare fraud (Marsh, 2020). Questions were raised in the UK Parliament around the use of the new housing algorithm. The new housing algorithm was stated to potentially further the UK

North-South divide, a divide exacerbated by the COVID19 crisis, which resulted in the "revolt of the Northern Metro Mayors," led by Andy Burnham, Mayor of Greater Manchester. The housing approach, including the use of the algorithm, was consequently dropped by the government.

The A Level fiasco happened against the backdrop of the Global Black Lives Matter movement and protests on the streets of the UK against structural racism and societal bias.

It is important to note that navigation of the use of data and AI in public services is an issue for all governments, not just one to be tackled by the UK alone. However, with the topic fully raised in the UK, the public and press are aware, there is a cause to look at the country as a case study. In this sense, review the current regulation, legislation, and governance structures and, in doing so, assess where the country is now and where the country is going concerning ethical outcomes of the implementation and use of data and AI. Is UK society really on the track when it comes to ensuring "tech is used for good?"

Against this backdrop, the objective of the study is to answer the following questions:

1. What are the key ethical issues raised concerning the implementation and use of data and AI?
2. How effective are regulation, legislation, and good governance in addressing the critical ethical issues associated with use of data and artificial intelligence?
3. How robust is the current UK system concerning the ethical use of data and artificial intelligence?
4. Where gaps exist in ethical governance and guidance, what are the possible immediate solutions available?

Limitations of the Research

The research will not review the trade-off between legislation, regulation, governance, and progress in the ability to innovate and develop in relation to data and AI use. The area of technology moves fast, and constant innovations enter the marketplace every day. In this sense, the research will remain a snapshot of time acting as a benchmark for progress in data, AI and ethics, specifically concerning the UK regulation and governance mechanisms.

Chapter 1

Methodology and Approach

1.1 Unstructured Interviews

Firstly, several unstructured interviews with professionals in central and local government were had. These were qualitative, allowing individuals to speak honestly and openly concerning data, digital transformation, and public services. The interviews helped assess the current internal landscape and critical departments within the UK government concerning data, ethics, governance, and investment concerning data and AI. Discussions helped to understand some of the essential barriers experienced concerning the digital transformation of public services, the future trajectory, and strategy. This research focuses on the legislation, regulation, guidance, and governance at a central level, devised to support all public institutions. At this stage of the study, the inclusion of local government helped inform the issues faced by those responsible for the direct delivery of public services. For example, the impact of austerity measures on digital transformation and how impactful current guidance around the ethical use of data and AI was in terms of implementation. This approach ensured that the research would be addressing live issues and adding to the current body of activity and work taking place with ethics, data, and AI use.

Finally, the unstructured interviews helped to focus the scope of the wider literature review of ethics, data, and AI in terms of relevance to the case study of the UK.

1.2 A History of Development

The first stage of the literature review was to summarise the history of technological revolutions and encountered issues. This section is followed by an in-depth review of the development journey of AI. The development journey of AI helped frame the timeframe of technological advancements (the speed) and the emergence of the literature concerning ethics and AI. The inclusion of the history of the development of AI provides scope in

terms of breadth of academic disciplines required for wider literature review on ethics and AI.

1.3 Ethics and AI Debate Areas

The initial scope of the ethics and AI literature review was informed by the unstructured interviews and the history of the development of AI, as stated above. Sources from across academic disciplines, including philosophy, behavioural economics, data and computing sciences, commerce, and business, were included. The literature review of the debate areas provided scope for the research to review and include the key ethical issues and questions from a global perspective. This global lens enables the presentation of various examples where ethics and AI have been dissected and tackled.

1.4 UK Case Study

The research focuses on the UK regarding regulatory, legislative, and governance frameworks currently in existence with regard to data, artificial intelligence, and ethics. Focus is given to the key bodies and institutions raised during the preliminary unstructured interviews at the start of the research. An overview of public sector values in the UK is provided. These values form part of the governance structure of all public institutions and thus inform use of data, ethics, and AI within public services. At the time of writing, a review had been recently completed by the UK Government into bias within algorithmic decision making. A synthesis of this review is included. The synthesis focuses the research on the key debate area of bias in decision-making and fairness.

1.5 Addressing the Gaps

One of the debate issues raised in the wider literature review and the UK case study was bias in algorithmic decision making. The final part of the research focuses on this debate area and gaps in governance.

The research conducted into the UK public sector raised the awareness of governance through values and the legal obligation of public services to proactively improve equality for marginalised groups. In this sense, there is an alignment with the nonprofit sector,

regulated by the UK Charity Commission. Charities registered with the Commission also have a legal obligation to improve societal issues experienced by marginalised groups. Due to this alignment in objectives and values, a focus is then placed on registered charities in the UK.

Unstructured interviews were held with individuals from registered charities in the UK. Discussions focused upon techniques, methods, and processes used within the sector to effect societal change, specifically concerning marginalised groups and discrimination faced by such groups. These unstructured interviews informed the focus for the final part of the research. While many techniques in the sector exist, the most popular for affecting and evaluating system and societal change was using the Theory of Change (ToC). Furthermore, the ToC was highlighted as facilitating partnership working and collaboration across industry sectors and other charities.

The research focuses on the ToC with regard to its history, evolution, and structure as a process within the nonprofit sector. This research resulted in assessing the potential role of the nonprofit sector in the UK in providing a solution for the highlighted gaps in the ethical governance and use of data and AI in the delivery of public services.

List of Key Definitions

Algorithm: Rules given to an artificial intelligence programme to help it learn on its own. Algorithms are essentially what makes a system intelligent.

Anticipatory Governance: Anticipatory Governance: The use of data to review behaviours and events in a predictable manner. The term also refers to decision making and the use of predictive methods to anticipate potential outcomes. It is a governance approach that aims to decrease risk and address such risks early on or prevent them entirely.

Artificial Intelligence: Theories and techniques which allow computer systems to perform tasks normally requiring human intelligence. It refers to a number of techniques called machine learning. All machine learning is AI. Not all AI is machine learning.

Automation: The production of goods and services with minimal human supervision.

Charity Commission: Non-ministerial government department in England and Wales which regulates registered charities.

Deep Learning: Models the brain, not the world. Networks of artificial neurons process input data to extract features and optimise variables relevant to a problem. As with Machine Learning, results are predicted to improve through training. Deep learning is therefore viewed as a subset of Machine Learning.

Digital transformation (DT): The ability to leverage technology and data impacts all aspects of culture, society, business and services and results in DT. DT is a transformation of operations, products, process and organisational structures, which results in doing things in a new digital way.

Diversity: Having differences within an organisation or setting such as identities, backgrounds and experiences.

Equity: Treatment of people in ways to make sure that they are not prevented from accessing opportunities and resources and to ensure that others do not have an unfair advantage. Essentially, equity means to offer individuals what they require to ensure fair access to opportunity and resources.

Inclusive Growth: Economic growth that is distributed fairly across society and creates and opens up opportunities for all individuals and groups.

Inclusion: A proactive approach to ensuring diversity which ensures that all people from all backgrounds, identities and experiences feel able to fully participate. The result is a culture within which individuals are enabled to be their whole selves.

Inequality: Economic inequality refers to the gap between the richest and the poorest, a gap which is growing (ITU 2020). Those at the top, the richest, get increased opportunities, influence, and power. Those at the bottom, the poorest, miss out on the healthcare and schooling they need. To tackle inequality, it is important to address influence, opportunity and deliver fairness for everyone. In this sense, inequality is a structural issue within our global society. Therefore, when looking to assess whether digital transformation is overcoming inequality, it is important to review just how this can or could undermine the existing structural inequalities that exist.

Internet of Things: A system of interrelated, internet-connected objects that are able to collect and transfer data without human intervention.

Intersectionality: An approach to equity, diversity and inclusion that understands that accounts for multiple forms of discrimination which are experienced simultaneously by individuals and groups. Multiple forms of discrimination can be, but are not limited to, disability, gender, race, religion, sexual orientation and socio-economic standing.

Logframes: A planning approach developed in the 1960s. The tool is a matrix that provides an overview of a project's goal, activities and anticipated results.

Machine Learning: This enables programs to learn through training instead of being programmed with rules. Through the use of training data, such systems provide results that are predicted to improve with experience over time. The aim of machine learning is to create systems such as algorithms that learn the relationship between input data and the actions that are wished to happen without being programmed.

Modern Technical Infrastructure: For the purpose of this research, modern technical infrastructure refers to the hardware, software, services, and security that enables the storage, transfer and processing of information and data, which in turn allows individuals and entities to interact in the digital space. Examples include the Internet of Things (IoT), cloud computing, next-generation wireless networks such as 5G, big data analytics,

blockchain, artificial intelligence (AI) and computing power. This is represented by the Information and Communication Technology sector (ICT).

Protected Characteristics: The groups of whom it is against the law to discriminate against. The classifications of protected characteristics are; age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex and sexual orientation. The Equality Act 2010 protects from these types of discrimination.

Smart Cities: An urban area that uses different methods to collect data. Methods include the use of sensors to collect data. The data is then used to manage resources, services and assets more efficiently.

Social Justice: Social justice is how fairness manifests in society. Social Justice is dependent and built upon four key principles. These principles are human rights, access, participation and equity. To achieve social justice is to achieve all four principles.

Chapter 2

The History of AI - Multidisciplinary Approach Combined with Culture

Today the global economy is rapidly transforming—the latest iteration of change, the implementation of AI into digital transformation has excellent potential for both industry and government, and society. Already industry sectors are witnessing rapid change as a result of the use of data and AI. For example, Fintech, the digitisation of financial services and capital markets. Also, in retail, where AI is used to provide insights on consumption and trends, through to the automation of warehouse operations. In retail, these are just a few processes that have benefited from the introduction of AI and use of data. Public services are also witnessing transformation, for example, healthcare and education.

The history of AI spans many disciplines and centuries. In this sense, AI, or at least its potential, has been in the human psyche for some time.

When reviewing what it means to be human, philosophy will often use machine intelligence. For example, philosopher Étienne Bonnot de Condillac in *Traité des sensations*, published in 1754, uses the parable of a statue and fills the statue with knowledge and senses to demonstrate and question at what point it becomes human. Science fiction writers have used intelligent robots in their writing. For example, L. Frank Baum in *The Wizard of Oz* and the description of a mechanical man in *TikTok*, written in 1907. There is also Mary Shelley's *Frankenstein*, written in 1818. In more modern times, the mode of the film is used. Talking robots that are human-like in Scott Lucas' *Star Wars* are just one famous example. Such philosophers, writers, and creatives have no doubt inspired AI researchers. However, they also succeed in capturing society's imagination, and in some instances, play into the fears of what might be.

It is possible, for example, to link the role of chess within AI back to the 18th century. During this time, it was popular to use clockwork mechanics to fool people into thinking a machine was playing chess. In 1997, one of the key milestones in the development of AI saw IBM's Deep Blue, a chess-playing computer, win a chess match against a reigning champion.

AI is not simply about the mechanics of a moving machine. AI is also about understanding the nature of intelligence. Many of the key figures in the development of AI have a varied background in terms of expertise. Annex 1 demonstrates the key milestones in terms of AI starting in the post-war era with the rise of modern computers. Key figures and disciplines include; Alan Turing (mathematician, computer scientist, logician, cryptanalyst, philosopher, and theoretical biologist); John McCarthy (computer and cognitive scientist); Herbert Simon (economist); Judea Pearl (Computer scientist and philosopher); Rollo Carpenter, creator of chatbots Jabberwacky and Cleverbot (programmer and inventor). Also included in the infographic are many companies, including; IBM, Sony, Amazon, Google, Alibaba, Microsoft, Samsung, and Facebook. Philosophers, authors, scientists, mathematicians, programmers, theorists, movie directors, and businesses have played a part in taking AI from science fiction to reality.

This multidisciplinary approach, combined with culture and the ability of AI to fundamentally change society, has given rise to literature in the discipline of Ethics of AI and Robotics.

2.1 Rage Against the Machine

The progress of humanity will always give rise to fears, and change is often something humans and society pushes back on. For example, during the late 1800s, during the Industrial Revolution, the then termed Luddites were a violent force against change, specifically within the textile industry in the UK. Groups of highly skilled artisan craftsmen would smash up machinery, afraid that their jobs were under threat. A reaction in part to

industrial responses to technology implementation. Namely, lower wages and hiring unskilled staff to work in the factories. Jobs became dangerous, hours worked became long, and skills learned through apprenticeships and lifetimes were junked. Perhaps if the so-called Luddites could share the gains from the new technology introduced in the industrial revolution, such social upheavals could have been avoided.

In Britain, more soldiers were used to prevent the violence caused by the Luddites than were used by Wellington to fight Napoleon on the Iberian Peninsula; people as young as 16 were hanged, and many others were sent off into exile to Australia (Klein, 2021).



A rioting mob of Luddites, British workers who were opposed to increasing mechanization of jobs. ime Life Pictures/Mansell/The LIFE Picture Collection/Getty Images

In summary, the historical consensus is that the Luddites did not fight against the rise of technology but a decrease in living and working standards; essentially, a loss of livelihoods and the rise of long and dangerous work in factories based in overcrowded urban centres. It was the implementation of the technology which caused such social upheaval, not the technology itself.

Concerns of new technology are not something simply confined to the history books of the Industrial Revolution. Some concerns are wrong, such as those surrounding the introduction of the telephone and fears of this destroying personal communication and interaction. Other concerns, such as cars killing children and changing landscapes, are correct, concerning, and very relevant. Other concerns such as digital photography and music-making vinyl records and photographic film redundant are relatively correct yet with moderate impact and relevance.

To avoid extreme and perhaps as has historically been witnessed, violent social upheaval and allay fears, rational or otherwise, it is critical to review the ethical aspect of implementing new technology. In doing so, it is possible to ensure new technology and innovation benefits society, leaves no one behind, and, where possible, changes existing unfair systems for the better.

Concerning technology, there has been no more an appropriate time to assess ethics than when considering the implementation and the rise of AI and Robots.

As discussed in the previous section, AI and Robots have been in the human psyche and have been for hundreds of years. Fears and moral panic, relevant or irrational, will exist. Furthermore, for humanity to realise the potential of AI and Robots for the common good, it is critical to review the ethical debate to ensure good and fair implementation.

While unethical and bad implementation of technologies today will unlikely result in hangings and exile for those who oppose it, as was witnessed during the Industrial Revolution in Britain, social unrest can destabilise progress and democracies.

The following chapter will review the academic literature that relates to key areas of ethical debate when considering AI and Robotic systems that can be autonomous.

Chapter 3

Literature Review - The Key Debate Areas

The following chapter will review the key areas of debate in the literature with regard to AI and autonomous systems and ethics. The debate areas chosen to be covered are those which are most relevant to the risks and barriers being faced by the public sector in the UK today. In this sense, the areas covered are by no means a representation of all the academic literature; for example, the area surrounding robots and robot rights has been excluded due to the lack of relevance at the time of writing of this debate area in relation to public services and digital transformation.

Some of the key debate areas overlap, for example, decision bias and transparent decision making. However, current legislation and regulation impact differently within the two issue areas. It is therefore important to be able to separate the issue areas to compare how each has and can be addressed in the case study of the UK in the next chapter.

3.1 Privacy and Surveillance

Privacy is a human right recognised in the U.N. Declaration of Human Rights. The right to individual privacy features in Article 8 of the European Charter of Fundamental Rights. Privacy links to values of freedom of association, speech, and dignity, and the right to privacy is in many global and regional treaties. In this sense, data and ownership of data belonging to the individual are linked to and embedded within the human rights agenda.

Data collection and storage today are rapid. Data collected includes sensor data, which offers a picture of an individual's offline life. Sensor data is accessed through technology such as wearable tech. Wearable tech can monitor human posture, heart rate, and other vital aspects. Sensor data can also be in the form of GPS, temperature, and motion tracking. Data surrounding an individuals' on and offline life can be captured, stored. The use of AI offers the opportunity to analyse this data and, for example, make predictions regarding

individuals in all aspects of life. This type of data collection and analysis provides a complete picture of an individual and impinges upon privacy. The capture, storage, and analysis of such a complete picture of data can be referred to as surveillance. The organisations that rely on capturing this data for their business models are stated to be the surveillance economy. Schneier writes, "Surveillance is the business model of the Internet" (Schneier, 2016).

Data collection is far from transparent and tends to be dominated by a few large tech companies. For companies such as Facebook, Amazon, Microsoft, Google, and Apple, the business model is to maintain users' attention and, in doing so, collect further data. Shoshana Zuboff, in her book *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, encourages individuals to consider the amounts of data fed to companies to answer three recurring questions, 'Who knows? Who decides? Who decides who decides?' (Zuboff, 2020). Answering such questions are key to politics and governance, and specifically within democracies. For democracy, it is, of course, the people and civil society who must know, decide, and decide who decides (Whitcomb, 2020). With the rise of the Internet of Things (IoT) and smart cities, the ability to access constant and real-time data on individuals at all times will only increase. This situation could lead to algorithms knowing us better than we know ourselves, as stated by Professor Harari in his book *Homo Deus: A Brief History of Tomorrow*.

Privacy-preserving techniques are available to tackle the issue. Privacy-preserving techniques can conceal the identities of individuals. Techniques such as randomisation, k-anonymity, and i-diversity can reduce data granularity and anonymise individuals and groups within the data sets. However, data is increasingly multidimensional, limiting the effectiveness of some privacy-preserving techniques while also driving down the quality of data (Chauhan, 2013).

Currently, privacy laws surrounding data privacy are being tested in the law courts. Unlike other civil liberties and property protection rights, those around data remain in their

infancy. The infancy of the law means that many firms operate without fear of impunity, and others, looking to implement AI may operate with extreme caution and risk averseness, which can slow progress.

One example of the wide-reaching implications of data privacy, regulation, and the law courts is the Schrems II case, Data Protection Commissioner v Facebook Ireland Limited, Maximillian Schrems ("Schrems II"; case C-311/18). Here, the Court of Justice of the European Union ruled that the Privacy Shield, the mechanism previously used to allow data transfer between the U.S. and the E.U, was no longer compatible with the EU General Data Protection Regulation (GDPR). A key issue in the Schrems II case was the access and use of data by the U.S. government in relation to surveillance for national security purposes. In essence, U.S. surveillance law was deemed to violate the privacy of E.U citizens and thus be at loggerheads with the EU Charter of Fundamental Rights. Although the ruling did not stop all transatlantic data flows, it prevented U.S. companies from holding or storing personal information about E.U. citizens, which has impacted commerce.

Cases such as GDPR and the Schrems II case demonstrate that intervention and legislation are operating within the area of data and data privacy. Therefore society is not following a path that believes technological solutions will solve societal problems without due guidance and laws.

3.2 Changing and Directing Behaviours

The use of data can result in undermining freedom of choice. However, manipulation of choice is nothing new. For decades marketers and advertisers have used various techniques to influence consumer behaviour, and in doing so, maximise profit.

One known approach used in commercial and public policy is "nudging." When considering public policy, many methods (such as taxation and legislation) are implemented to achieve policy outcomes—as such, using policy to change behaviour is nothing new.

Nudging, which rose to prominence in the UK during the Cameron years' is considered a form of "libertarian paternalism." The phrase liberal paternalism was coined in the book *Nudge: Improving Decisions About Health, Wealth And Happiness*, co-authored by Thaler and Sunstein. Thaler and Sunstein build on the work done by Daniel Kahneman in his book *Thinking Fast and Slow*. In summary, Kahneman states that humans have two systems of thought. These two systems of thought are called the dual processing theory. System 1 uses heuristics, a form of mental shortcut, to find a simple answer to a complex problem. System 2 is activated during complex thinking, thus, thinking slow. Because system 1 is quick, it tends to form the basis of human response, yet this system can often be wrong due to heuristics and environmental factors. Thaler and Sunstein state that Nudge appeals to System 1, the more dominant system, and through a variety of measures, "nudges us" to make the correct choice, and this is a form of libertarian paternalism. For example, by banning smoking in public spaces, policymakers are nudging in the sense that smoking starts to become less socially acceptable. However, there is a fine line between nudging people's behaviour and changing people's behaviour.

In the Facebook Cambridge Analytica scandal, data and mining of it to influence voting patterns were exposed. Psychological profiles of social media users were built using data. Targeted adverts used the psychological profiles; adverts were tailored toward an individuals' fears, emotions, and needs to nudge and influence choice during elections. It was also revealed that data gathering happened without the prior consent of individuals. The result was claims of manipulation of free and fair elections.

While the E.U has strengthened privacy protections with the GDPR other countries have not because of the trade-off. Increased protections and regulations can slow down progress and innovation, and thus competitive advantage. AI technology can invade privacy and is proven to be very adept at manipulating people, as discussed in seminal documentaries such as *The Great Hack* (Amer and Noujaim, 2019) and *the Social Dilemma* (Orlowski, 2020).

3.3 Transparent Decision Making

Automated decision-making and processes make it difficult for individuals to understand how a system comes to decide what it does. This inability to identify patterns in the decision-making process causes issues when looking to address bias. Where machine learning is involved, such decisions and patterns will also be opaque to the expert. As such, opacity and bias are linked, and both require a political response.

AI systems that rely on machine learning techniques extract patterns from datasets. Programmes evolve as new data is or feedback is entered, altering patterns used by the system. In this sense, the outcome is that the decision made is not transparent for the user or programmer. The programme is also reliant on the quality of data entered and the labels used. Therefore, if police add data, potentially, race bias will be reproduced if this data is inherently racially biased.

Many criticisms exist when considering transparency and the "black box" of deep learning. The "black box" refers to the artificial neuron networks. This is the piece of a computing system that simulates the way the human brain analyses and processes information. For example, someone receiving a medical diagnosis or a decision surrounding finance will want to know why a certain decision has been made (Marcus, 2018). In this sense, transparency and bias are raised as drawbacks for progressing AI. Geoffrey Hinton, the co-author of a paper which is now known as being central to the expansion of AI, calls for a new path to AI development. Hinton calls for the dispense of labelled data. For Hinton, neuron networks must get to the point of operating in the same manner as a human brain. This approach is termed "unsupervised learning" (LeVine, 2017), machines operating exactly as the human brain. However, unsupervised learning raises issues of accountability.

There are proposals for how to make AI more transparent. For example, Nicholas Diakopoulos discusses algorithm accountability. Diakopoulos's reference to the power structures and resultant biases sets a case for reverse engineering. Reverse engineering can deconstruct algorithm power structures and thus make them increasingly transparent.

For Diakopoulos, there are four power structures. *Power structures* are the steps taken to solve a problem. Firstly, prioritisation, the criteria used to provide a ranking. Secondly, there is classification, when the algorithm makes false positive or false negative classifications. For example, content may be classified as fair (false negative) when it is not. A fine-tuned algorithm increases the volume of false negatives, which are not fair, to be forwarded, and other fair decisions get increasingly turned down. Thirdly there is the association, the relationship between entities. Association builds a context around people; for example, an individual could be associated with a felon due to someone in their family having a criminal record. This association can result in an individual being turned down for services or highlighted as a risk by the association. Finally, there is filtering, the exclusion of information based on set criteria. Filtering can be used to sensor information or inadvertently create an "echo chamber" where users see only one type of content, thus stifling thought (Diakopoulos, 2014). The ability to dissect the listed power structures through transparency regulations is just one solution and one around the issues of the "black box" and opaqueness of decisions.

It is also possible to witness a shifting regulatory environment within the area of transparency. E.U regulation states that individuals have a right to an explanation concerning decisions made using data. (The European Parliament and The Council Of The European Union, 2016). This legislation, therefore, starts to shift the accountability for decisions to those implementing the technology.

While some argue that standards and explanations for machine-based decisions are higher than for those of humans, without transparency processes, it will be hard to judge if the social and environmental benefits such systems and machines can bring are achieved.

3.4 Decision Bias

AI uses data (the input) to make decisions (the output). Data can be used to make future predictions; for example, when assessing whether to give an individual a loan or a

mortgage, data will be used to determine creditworthiness. Such predictive situations raise a key question around bias in decision-making.

Most discussed is the issue of bias and predictive policing. While the use of algorithms in policing is patchy across areas, most common and assessed within the literature is the use of AI to assist with resource allocation; predicting, and pre-empting where crime is likely to occur. As noted in the paper *Algorithms, Human Decision-Making, and Predictive Policing* (Phillips and Pohl, 2021), bias within policing can result from the human decisions made regarding the construction of the algorithm. For example, areas are carved into blocks of space, and these blocks are ranked as part of the decision-making process. Various aspects such as both historic levels of crime and recentness of crime levels within the blocks are taken into consideration to predict where resources are required the most. Phillips and Pohl state the ranking of blocks causes issues because the prediction is simply that of an expected rise in crime within a block, not the spread of crime across a city. In this sense, the human construction of the predictive policing algorithm can result in bias. Phillips and Pohl refer to this as the problem of space delinearisation, which in turn results in prioritisation and ranking bias, as referenced earlier. In summary, heuristics exist at the time of the conception and creation of the algorithm, resulting in a bias. In this sense, how we choose to interpret the world in numbers is equally important as the data used; models matter.

The second risk to demonstrate is the implementation of AI predictions. Peter M. Asaro describes the creation of a Strategic Subject List for the Chicago Police Department (CPD). Forty-eight factors and data from individuals' arrest records were used to compile the list. The list included the risk of an individual being a victim of crime or violence. The list also considered an individual's social network; this included whom someone had been arrested together with. These and other factors were used to create a score. The score resulted in 1,400 individuals highlighted as "high-risk" of being involved in violence. This metric contained people who had never been arrested but were indeed victims or in the social networks of victims or perpetrators. Officers stated that they were misled and not informed that the list combined those at-risk of being victims and presumed all were potential

perpetrators. The result was harassment of individuals by the CPD. It is with this in mind that Asaro raises the important question of the Ethics of Care. Asaro writes that if the aim is to prevent crime, timely interventions for those "at-risk" should be used as part of this prevention instead of viewing predictions as Models of Threat (Asaro, 2019). In this sense, the effective implementation of models is also key to avoiding bias. The question arises, who is accountable for such implementation, those building models, those using the models, or both?

Bias surfaces when unfair judgments are made on irrelevant characteristics or a preconception. This behaviour is a learned bias. Humans are subject to cognitive bias and confirmation bias. In the example where predictive models are used to assist police in terms of resource allocation, user bias could arise in terms of confirmation bias. In this sense, bias can occur from the implementation of machine learning and algorithm models linked to user behaviour.

Algorithmic bias is also introduced through data. How data is collected, adapted, and entered. Furthermore, some algorithms collect their own data based on human-selected criteria. Such criteria can reflect the bias of human designers. One famous case of data bias is that of the Amazon AI recruitment tool. Amazon used employee data to train their AI model. With most technical positions filled by men, this societal bias became embedded in the model, and the model learned to penalise female candidates. Amazon subsequently stopped using the recruitment tool. However, it is a very good example of how societal biases in data can result in systems that discriminate on gender, race, disability, and other group categories.

Research into how to integrate debiasing capabilities into AI algorithms is underway. For example, in 2019, research was published on integrating debiasing capabilities directly into a model training process to uncover existing bias in training data (Amini et al., 2019). However, the notion of fairness limits the ability of technical solutions, with concern expressed over a "mathematisation of ethics" (Whittaker et al., 2018).

3.5 Employment and Automation

Automation in industry is nothing new. Automation plays a role in determining productivity and has done so historically. In the United States, in 1790, nearly 90% of the working population were farmers. By 2000, the population employed in farming was 1.9%. Despite the drop in labour numbers, production output has radically increased, resulting in lower prices and a labour force-free to conduct other tasks. The result of such productivity gains is economic growth and wealth creation.

The digital automation of today replaces certain tasks, information processing, and thought. In this sense, it is not driven by mechanical machines, as was witnessed in the first industrial revolution. This factor makes digital automation less costly and potentially more rapid. The result means very quick change within the labour market. Some argue that automation has meant a loss of shared growth and that the next phase, AI and machine learning, could exacerbate this trend.

In his article *Remaking The Post-COVID World*, Daron Acemoglu draws upon his research to present a case of excessive automation, with his focus on the U.S. He argues that excessive automation is that which is not thought through. He states that excessive automation fails to increase productivity and delivers high social costs concerning low wages and low employment, which fuels inequality. Behind this excessive drive of automation, the author references the monopolistic nature of Big Tech with business models centred around automation and algorithms. Acemoglu states that due to the dominance of Big Tech, diversity of perspective is lost. Diversity of perspective has been the anchor of successful technical innovations in the past. Drawing on research over the past 40 years, the author also references the U.S. tax system and the fact that capital continues to be taxed less than labour. This, Acemoglu states, creates an incentive for businesses to undertake excessive automation (Acemoglu, 2021).

Government intervention policies that accelerate investments into renewable technologies combined with increased minimum wages to support the demand side are advocated by both Larry Summers (Summers, 2016) and Acemoglu as solutions. Contrary to the excessive automation argument made by Acemoglu, Summers references the role of the "new economy," namely the Big Tech giants, as those who conserve capital and restrict investment in infrastructure; for example, the impact of Airbnb on hotel construction, Amazon on the construction of shopping malls and Uber's impact on automotive demand (Summers 2016). Summers also states that investment in new technology can be delayed or deferred for fear of it becoming obsolete. However, he references demand for printers, copiers, and office space instead of the continued iterative investment in automation technology referred to by Acemoglu in his article Remaking the Post-COVID World.

Summers and Acemoglu both state the need for regulation to prevent the unjust distribution of resources and wealth due to aspects such as the monopolistic tendencies of Big Tech firms. In this sense, the AI economy can be viewed as a driver of inequality if left unregulated. Regulation can also direct innovations within the AI economy into renewable energies, lowering the carbon footprint the industry generates as a cost to society.

In the radio podcast The Real Story: The Pandemic Brings More Robots, contributors emphasise that automation and AI are value-neutral, but it is how society prepares for and implements such technologies that are key to deciding whether they pose a threat or an opportunity. In this sense, it is how humans design the vision of "better" that matters. Presented is the scenario of a society where human interaction is reduced to zero at every turn. This scenario and the implications it presents should guide investment decisions in automation. Automation should always be thought through.

This scenario planning, in some ways, links back to the argument presented by Acemoglu, around unthought through excessive automation. Therefore the decisions made around where to use and invest in automation and AI created by the public and private sectors are of ethical importance. Contributor to the radio podcast, Rob Carpenter, CEO, and Founder

of Valiant AI, states that although the pandemic has increased the pace of automation and AI by approximately 5-10 years, key technologies such as driverless cars will continue to take decades to develop.

Carpenter states that, therefore, automation AI use and its implementation is slower than it feels. Carpenter also points out that the use of automation and AI can make jobs more manageable instead of replacing them. He states that this is especially true in sectors with staff shortages, such as the restaurant and hospitality sector, post the COVID pandemic. The result here is not a displacement of human labour but an improvement in working conditions by automating specific tasks instead of whole job roles. However, investment in machines and ranking within organisations has raised concerns. In Amazon, machines terminate employment based on a ranked performance criteria. This use of machines leaves the workforce with little room for discussion and people feeling fired from their jobs when they consider they have done very little wrong. The use of automation in this way again raises the question of where automation should take place and where it should not, where human supervision and decision-making should prevail within workforce management, and where it should not (Soper, 2021).

The final discussion of the podcast, *The Real Story*, centres around the experts emphasising the need to address that people must not be left behind due to automation and AI use. Post the COVID pandemic, many executives and CEOs may now be more open to introducing automation and the use of AI. Work will still exist for humans, where they must make judgements on nuanced aspects and complex situations, more in line with human traits than machines. The result may be that lower-wage jobs are displaced and highly skilled new jobs created. Therefore, there is a need to build a bridge for people to transition into high-skilled work which involves cognitive tasks. In this instance, what matters is not automation, and AI use itself, but who has routes into the new jobs created and how the education system prepares people. However, the education system is subject to gender, race, and disability inclusion issues. Currently, in the developed world, investment in education is made in the first 20 years of human life. This early investment in education

will need to transform into life-long learning, which needs to be enabled. For example, women, who are more at risk of losing their job due to automation, require support through the financial barriers of retraining and re-skilling.

In the podcast, the panel touches on how the productivity gains of automation and AI use can be shared. Such gains could include more time away from the office or the creation of citizen wealth funds. Furthermore, the introduction of increased value and thus wages for key jobs in the care sector, which are less likely to succumb to automation, can occur.

Finally, the issue of power is discussed. People who own technology have power. Suggested is the need to take collective action and secure an agreement on the future vision to address this. If workers lead automation in identifying opportunities and ensuring shared prosperity, protection of labour rights, and ladders between jobs, better outcomes will be realised (Henley, 2021).

In summary, concerning automation and AI use, much of the literature and discussion centres not on the technology itself but implementation, how society prepares to ensure no one is left behind, and how productivity prosperity is shared fairly.

3.6 Autonomous Systems

Concerning ethics when considering autonomous systems, the debate rests on who is in control and who is responsible. In this sense, assessing the power dynamic is similar to the issues raised within bias and opacity. The reason for this is because a system is only autonomous to a certain degree. Human control is still in existence.

One of the more debated areas of autonomous systems is that of cars. The reality of autonomous cars becoming common on the roads anytime soon is low. However, there is still a need to consider how responsibility and risk are allocated. The vast majority of ethical issues when considering autonomous vehicles are covered by legislation already.

For example, keeping a safe distance, speed limits, and so on. In this sense, programmers of vehicles need to simply ensure that the vehicles follow the rules as already laid down. This fact does, however, affect the current distribution of responsibility. For example, currently, the driver, the individual, is responsible for the driving and the manufacturer for the technical safety of the car. Policy efforts in this area have started to address this aspect, shifting accountability and responsibility from the motorist to the manufacturer and operator of the vehicle (Duisberg and Vogel, 2017). What is clear is that the use of autonomous cars could potentially lower road traffic accidents and deaths, making matters better, not worse, concerning driving.

The discussion of autonomous cars leads to the much-debated use of autonomous weapons. It is unclear whether their use could make the matter of war better or worse. Again, the issue tends to rest on responsibility, is the human or machine responsible? In his article *Killer Robots*, Robert Sparrow considers a situation where autonomous weapons are involved in a war crime. Sparrow argues that deploying weapons with sophisticated AI can only be justified if someone can be held accountable for the decisions such weapons make. It is not possible to punish a machine, and those who order their use will not have complete control over the decisions made. Sparrow concludes that the use of autonomous weapons is unfair to the potential casualties and the person responsible for authorising their use (Sparrow, 2007). In this sense, when considering the use of autonomous weapons, the ability to identify risk and decision-makers is critical, and it raises questions around, just because something can be done, should it indeed be done. In the case of autonomous weapons, 4502 AI and Robotics Researchers and 26215 other endorsers, including Elon Musk, Steve Wozniak, Noam Chomsky, Jack Dorsey, and Stephen Hawking, wrote an open letter denouncing the start of a military AI arms race and calling for a ban on offensive autonomous weapons beyond human control (AI & Robotics Researchers, 2015).

3.7 Superintelligence brings The Singularity

The emergence of superintelligent machines that humans can not predict gives rise to The Singularity debate. The Singularity is when computing power transforms human life into something that can not be recognised today.

Futurist Ray Kurzweil has claimed that The Singularity will happen by 2045. In this sense, AI will pass a valid Turing Test. The Turing test is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human (Reedy, 2017).

John R Searle uses the analogy of the Chinese Room to demonstrate the difference between simulation of and duplication of consciousness and states that this is a crucial point when considering implications of the potential of The Singularity. In brief, the Chinese Room argument is where a person is locked in a room unable to speak Chinese, only English. The person receives instructions in English that allow them to match Chinese characters to English words. This situation gives the impression that the person can understand Chinese when they can not. In this sense, concerning the Chinese Room, Searle demonstrates that computer programmes can simulate behaviours and interactions linked to human consciousness, not duplicate them and that this distinction is critical when considering future implications (Searle, 1980).

Others, such as Neuroscientist Christof Koch, state reproduction of the same kind of relevant relationships among all the relevant neurons in the brain will create consciousness. He states that the interconnectedness of the internet may already have some form of conscious state (Closer to the Truth, 2021).

Aside from the debate around when The Singularity will be reached and what defines the consciousness which enables it, there are those in the literature which warn of the risks and those which write of the benefits. For example, Nick Bostrom argues that the outcome of developing superintelligence is likely to be catastrophic with a need for care in system

design and governance. For Bostrom, the issue of superintelligence and its potential development is helpful in that it stimulates and helps humanity take a safe and beneficial approach to AI development today (Brundage, 2015). Ray Kurzweil, however, focuses on the benefits of the merging of humans and machines into a new form of non-biological intelligence that will solve such issues as pollution, hunger, and ageing (Kurzweil, 2006).

3.8 Summary

Except for autonomous weapons, a clear theme in the literature is that technological developments are relatively value-neutral. It is the design and implementation that poses benefits or risks to society and certain groups.

The literature demonstrates a need to define with whom responsibility rests when negative outcomes arise. When considering autonomous vehicles, the ethical matter of responsibility, where the responsibility for poor outcomes is transferred from the driver to the manufacturer, best demonstrates this issue. However, this becomes vague when considering other negative outcomes driven by ethical considerations such as opacity, bias, manipulation, privacy, and social issues arising from excessive automation.

The literature also raises the issue of fairness. How is this defined, and what are the parameters? Without clear definitions and parameters, how can the potential of unfairness be recognised, policed, and accountability for unfair outcomes assigned? Another question raised in the literature is whether the use of AI can and should be equipped with making better decisions than humans when concerning issues of bias.

The use of AI is linked to human rights. Demonstration of this is when considering privacy and surveillance and manipulation of decisions that can and have occurred through the use of data.

The issue of Singularity, while disputed to some degree, does achieve majority consensus that it is indeed possible and likely to happen. This debate, as stated by Bostrum, is most helpful to consider when looking at how to steer AI and technological developments to an ethical and beneficial outcome for humanity.

When considering the issue of autonomous weapons and indeed excessive automation, it is clear that there is a need to decide when not to design and, or implement technology and AI. Just because it can be done does not mean it should be done. A clear analysis of risk vs benefit combined with complimentary social solutions should be conducted to resolve this. For example, for automation, policies that support those at risk of being left behind need to be developed in tandem with the implementation of technology itself. Furthermore, existing bias and discrimination within supporting policy areas, for example, education, must be addressed too. Failure to do so may indeed risk societal blame levied toward AI and technology, failing to recognise the intersectionality which exists.

Accountability, privacy, fairness, explainability, and an alignment with the social norms of society are highlighted as key themes when considering AI Ethics. Trust risks being lost without key pillars such as these in place. If trust from civil society and individuals is lost, the industry may risk a return to an AI winter, as experienced in the 1980s. Decreased funding will, in turn, prevent progress in realising the benefits technology and AI can bring, especially concerning the Sustainable Development agenda and delivery of the United Nations Sustainable Development Goals (SDGs). In this sense, AI ethics plays a key and major role in ensuring how some of the world's most complex issues can be solved.

Good and quality guidance and regulation around AI Ethics can play an important role in preventing "moral panic" in society and guide design and implementation to better and improved outcomes for people and the planet while helping to avoid historical mistakes of the past.

Chapter 4

Ethics, Data and AI - The UK Case Study

4.1 Why the UK?

The following section will build upon the literature surrounding Ethics and AI to review current guidance, safeguards, and enforceable legislation surrounding the use of AI within the public sector.

As stated in the introduction and methodology, the UK is a case study due to several high-profile scandals that arose in the summer of 2020.

As demonstrated in the section reviewing the history of AI, notions of machines have been in the human psyche for hundreds of years. Many writers, poets, film directors, and philosophers have presented a dystopian vision for humanity and the future with machines. Such scandals involving technology, therefore, have the power to derail progress in the area of AI.

Scandals present checks and balances for the use of technologies too. In this sense, a high-profile scandal can not, therefore, be seen as terrible either. Specifically, the A-Level fiasco in the UK, where an algorithm used to predict exam results for students, which were unanimously considered unfair, raised the profile and need for a focus on data, AI and Ethics within the public sector. While citizens choose whether they use a private service or not, this is not always the case with public services. So, a research focus on public sector governance structures in relation to the use of data and AI in this field is relevant due to the societal impacts which occur alongside implementation.

Furthermore, a reason for reviewing the UK as a case study is that the government has allocated significant resources to this technology through the AI Sector Deal. The £1Bn investment put forward by the UK Government does not match the £20Bn invested by the European Commission; however, the UK investment has resulted in a focus on AI and Ethics.

4.2 The Nolan Principles - Towards Values-Based Governance

In the UK, anyone who works in public office or any private sector firm delivering public services must adhere to the Seven Principles of Public Life. These seven principles are known as the Nolan Principles. The Nolan Principles are mentioned in Ministerial Codes, form the basis of the code of conduct for local authorities, and are incorporated in the documentation for public sector organisations. The Nolan Principles are essentially a form of good and ethical governance driven by a focus upon behaviours and culture.

Therefore, the UK must ensure that the principles are upheld when developing and implementing the use of data and AI.

The Seven Principles of Public Life are:

Selflessness: Holders of public office should act solely in terms of the public interest.

Integrity: Holders of public office must avoid placing themselves under any obligation to people or organisations that might try inappropriately to influence them in their work. They should not act or make decisions to gain financial or other material benefits for themselves, their family, or their friends. They must declare and resolve any interests and relationships.

Objectivity: Holders of public office must act and take decisions impartially, fairly, and on merit, using the best evidence and without discrimination or bias.

Accountability: Holders of public office are accountable to the public for their decisions and actions and must submit themselves to the scrutiny necessary to ensure this.

Openness: Holders of public office should act and take decisions in an open and transparent manner. Information should not be withheld from the public unless there are explicit and lawful reasons for so doing.

Honesty: Holders of public office should be truthful.

Leadership: Holders of public office should exhibit these principles in their own behaviour. They should actively promote and robustly support the principles and be willing to challenge poor behaviour wherever it occurs (Committee on Standards in Public Life, 1995).

When considering the key issues raised in the literature review of this thesis, it is clear that the development and implementation of AI and new technologies within the public sector could indeed lead to a breach of some, if not all, of the Nolan Principles. With reference to issues raised earlier in the literature review, the principles of openness, accountability, and objectivity are notable. It is also true that the development and implementation of AI and new technologies could enhance the delivery of the Nolan Principles due to the technology itself being "value-neutral."

Reasons for introducing the Nolan Principles 25 years ago centre on the need to maintain public trust and thus legitimacy. To breach the principles is equivalent to breaking public trust and confidence and thus undermining the legitimacy of the public sector and its democratic foundations. To enhance the delivery of the principles will be to rebuild confidence and public trust and protect democracy.

In this sense, robust guidance around data, AI and ethics is crucial and not a "nice to have" within the public sector. The introduction of new technologies must be primarily values-based, and values tested instead of productivity and efficiency-driven. Indeed, it is a values-based approach that results in a good governance structure. For example, within a good values-based governance structure, the spending of public money is wise, hence adopting the principles 25 years ago.

Finally, robust guidance and regulatory framework about data, AI and ethics concerning its use in the public sector, which aligns with the Nolan Principles, will ensure better and more increased uptake of AI and new technologies. A good approach to AI and ethics will also ensure a good, if not changed relationship, is maintained between UK citizens and the state. In this sense, standards and governance are not barriers to innovation; when translated to the implementation of ethical standards, they will likely result in an accelerated use of data and AI.

4.3 The Current Ecosystem - Key Institutions in the UK

The 2018 AI Sector Deal led to the creation of the following institutions.

A government office for AI - The Office for Artificial Intelligence is a joint unit across the Department for Business, Energy & Industrial Strategy (BEIS) and The Department for Digital, Culture, Media & Sport (DCMS). It is responsible for overseeing the implementation of the AI and Data Grand Challenge¹. The Office for Artificial Intelligence recently published guides on using AI in the public sector and AI procurement guidelines. The office also published the Ethics, Transparency and Accountability Framework for Automated Decision-Making, aimed at civil servants.

An industry-led AI Council - The AI Council is an independent committee of members from academia, government, and the private sector. The council's role is to support the growth of AI in the UK and promote its adoption. The AI Council advises the Office for Artificial Intelligence. It has key aims, including developing public understanding and tackling negative perceptions, increasing skills and diversity of people working and studying AI, developing safe, fair, legal, and ethical data sharing frameworks.

¹ The Grand Challenges were part of the UK 2017 Industrial Strategy, Building a Britain fit for the Future. At the time of writing the Industrial Strategy is in the process of being transferred to the Levelling Up programme within the Treasury. Despite this, many AI Challenges have already been funded and are running.

The Centre for Data Ethics and Innovation (CDEI) - is an independent body that advises the government on AI and data-driven technologies. The CDEI develops regulation and governance. It provides expert advice on the ethical and innovative deployment of data and AI. The CDEI is still in a pre-statutory phase. Its remit is both the public and private sectors.

These three bodies work alongside the Alan Turing Institute. The Alan Turing Institute was made the national institute for artificial intelligence and data science by the government in 2017. The Government Digital Service (GDS) also plays a role in AI policy. Based in the Cabinet Office, in the heart of central government, the GDS is responsible for digital transformation within government. The GDS sets and enforces standards for digital technology, including procurement. The GDS currently works with a limited number of local authorities. The GDS aims to mask the complexities of public services for the user. The aim is to create a joined-up, personalised and proactive approach for all public services across the lifetime of a user. For example, services will range from registering a birth to reminders for the Ministry of Transport Test (MOT) for a car.

4.4 Ethical Guidelines for Data and AI Use in the UK

It is understood in government that the use of AI requires an adaptation of existing governance and management structures for the use of data and AI. The Office for AI, the GDS, and the Turing Institute published A guide to Using Artificial Intelligence in the Public Sector (Central Digital and Data Office and Office for Artificial Intelligence, 2019).

Within this guide is detailed guidance across the use of all aspects of AI implementation. Primarily the guide covers how to assess, plan and manage artificial intelligence; using artificial intelligence ethically and safely. Concerning ethics and safety, the guide includes a summary of the more detailed guidance compiled by The Alan Turing Institute. The Information Commissioner's Office (ICO) has published guidance on data protection compliance for the use of AI. The guidance is not a statutory requirement but a framework

for good practice to mitigate risk and ensure compliance within the existing legal framework. The guidance is focused on data protection compliance only and does not provide ethical principles for the use of AI. However, as demonstrated in the literature review, data protection and privacy overlap with the ethics debate, specifically concerning human rights.

The Office for AI has produced guidelines for procurement. The guidelines were developed with the World Economic Forum Centre for the Fourth Industrial Revolution, GDS, Government Commercial Function, and Crown Commercial Service. The ability to procure technologies is an enabler for the adoption of AI. The guidelines come complete with a toolkit, co-created with the World Economic Forum.

The detailed ethical guidelines from the Turing Institute are an attempt to create a shared culture of values. Firstly, the SUM values. These values incorporate elements from bioethics² and human rights³. They are guiding values for use throughout the innovation life cycle. The SUM values are below.



SUM Values (Leslie, 2019, p. 9)

² The principles of bioethics include respecting the autonomy of the individual, protecting people from harm, looking after the well-being of others, and treating all individuals equitably and justly (Leslie, 2019)

³ The main tenets of human rights include the entitlement to equal freedom and dignity under the law, the protection of civil, political, and social rights, the universal recognition of personhood, and the right to free and unencumbered participation in the life of the community (Leslie 2019).

The SUM values are designed for use when considering whether the AI project is ethically permissible. The values should be used by professionals when considering whether to use AI or not. In relation to the design and use of an AI project, the guidelines also include the FAST principles.

The FAST Track principles fill the void of accountability. The FAST Track principles aim to create a responsible and ethical environment for data innovation. The acronym FAST stands for; fairness, accountability, sustainability, and transparency. The principles are shown in detail below.

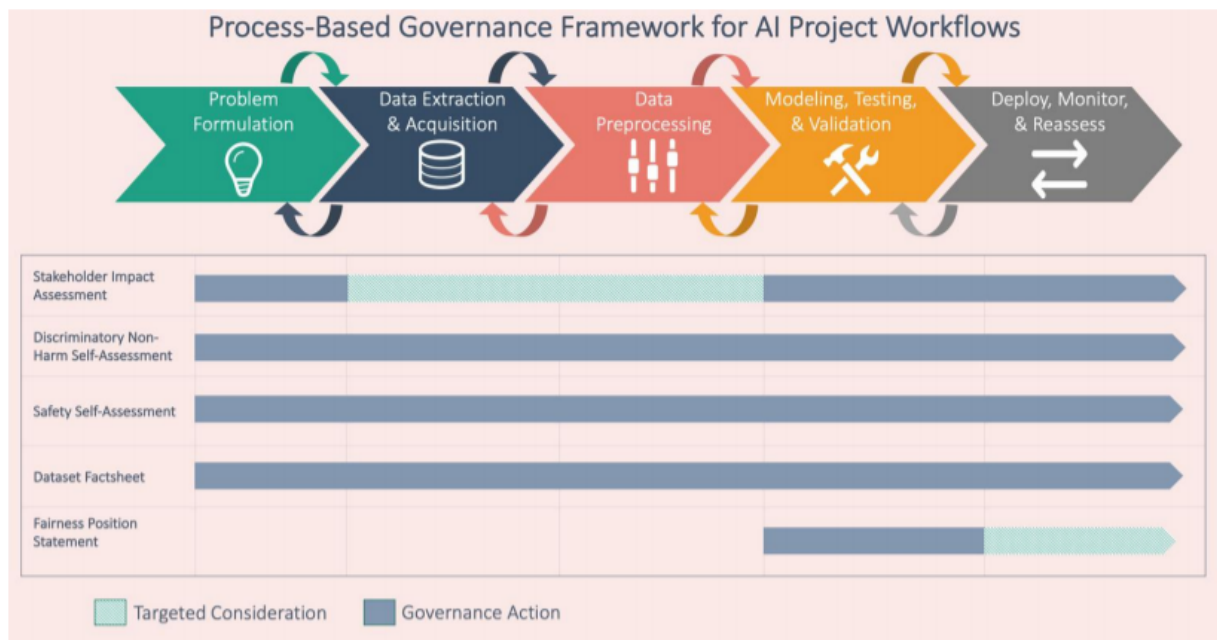


FAST Track Principles (Leslie, 2019, p. 12)

The principles of transparency, accountability, and fairness link to data protection principles when personal data is used. In this sense, the FAST Track principles are linked not simply to ethical guidance but legal requirements as laid out by GDPR and the Data Protection Act of 2018 (Data Protection Act, 2018). Therefore, the FAST Track principles link directly

with the ICO guidance on data protection compliance for the use of AI and include aspects of statutory and legal regulation, in addition to ethical guidance.

Finally, the guidelines demonstrate how to operationalise the SUM values and FAST Track principles through the use of a process-based governance (PBG) framework. The framework organises the governance considerations and actions to enable this.



Process-Based Governance Framework (Leslie, 2019, p. 38)

The SUM Values and the FAST Track principles are only one set of ethical guidelines advocated by the UK Government. Two additional sets of guidance include; the OECD AI Principles⁴ and the Data Ethics Framework published by DCMS⁵. Guidance on how all work together is not provided. This lack of clarity creates the potential for confusion in this space, both for the public and those looking to implement AI. The guidance provided is not sector-specific. As referenced when considering the history of AI earlier, the impact and

⁴ (OECD, 2021)

⁵ (DCMS, 2020)

relevance of the introduction of technologies vary. Introduction of AI within healthcare, policing and the judicial system may have increased impact and thus relevance in terms of ethical concerns than introducing similar technologies into the transport system, for example, managing traffic lights for better traffic flow.

4.5 Data, AI, Ethics and the UK Rule of Law

While data and AI ethics concerns mitigating potential future risks, it also operates within the existing legal framework. Furthermore, legislation and regulation are now starting to catch up with technological advancements. Without a good understanding of the existing regulatory environment, advancements in the adoption of AI and technologies may be lost and benefits to be gained missed. The key relevant legislation includes:

The General Data Protection Regulation 2018 (GDPR)

The GDPR has direct application in UK law through the Data Protection Act 2018. In the UK, the ICO is responsible for enforcing the GDPR. Penalty fines of up to £20 million or 4% of annual turnover can be levied for those who break the law. The GDPR covers key ethical issues concerning openness, responsibility, and accountability.

- **Openness:** Organisations must inform individuals before processing their data. Provisions are made within GDPR in terms of automated decision-making and profiling. If automated decisions are made about people, individuals must know what information was used, the relevance of the information used, and the likely impact. The GDPR makes provisions for individuals to question the processing of their data after decisions have taken place.
- **Responsibility:** Article 22 within the GDPR prevents the removal of human responsibility in the decision-making process. This aspect of responsibility is bolstered by ICO guidance. In ICO guidance, it is made clear that a public official automatically approving an AI decision does not create enough human involvement in the decision-making process.

- **Explanation:** While the right to explanation is uncertain within the GDPR, administrative law and the right to an appeal in UK law create a legal incentive to explain any public sector decision. Work is currently underway to clarify legal aspects around the right to explanation. This workaround, the right to explanation, is being conducted jointly by the Alan Turing Institute and the ICO (Covell 2019).
- **Accountability:** The GDPR covers accountability and states that organisations must document processing activities involving data and carry out data protection impact assessments. In this sense, any AI that processes personal data must comply with such requirements. This aspect of the GDPR means a regulatory environment that places accountability at the door of the organisation (The European Parliament and The Council Of The European Union, 2016).

The Equality Act 2010

The Equality Act 2010 is enforced by the Equality and Human Rights Commission (EHRC). The Act prohibits discrimination against certain protected characteristics; however, the Act does not consider socio-economic status. The Act also established the Public Sector Equality Duty (PSED) in 2011. The PSED ensures that public bodies take a proactive approach to fighting inequality. In reference to the earlier literature review, the establishment of the PSED can be said to rule out the option of AI maintaining the "status quo" concerning bias; this is because the PSED requires organisations to advance equality (Equality and Human Rights Commission 2011). However, there is no specific guidance for compliance with the Equality Act 2010 when considering the introduction of AI systems and automated decision-making. Key legal issues remain unanswered by the current legal framework when considering bias. For example, how is discrimination detected? What types of algorithmic profiling are discriminatory? Guidance should also go beyond legal compliance, addressing how to highlight and mitigate bias and when bias needs to remain.

4.6 Review into Bias in Algorithmic Decision Making

In November 2020, the CDEI conducted a review into bias and algorithmic decision-making. The report was timely. The UK had witnessed in the streets protests from

the Black Lives Matter movement; the COVID-19 pandemic had cruelly exposed painful realities around inequality and the exam results fiasco exploded. Public and press attention was firmly focused on the issue of fairness, and algorithmic decision-making was certainly receiving bad press.

With an acknowledgement that data can indeed make matters of inequality and unfairness worse, the CDEI conducted a sectoral review on algorithmic bias.

The issue of fairness and bias concerning AI and algorithmic decision-making is one of the weaker areas in terms of having robust regulation. Arguably the most robust regulation is that of the GDPR, which focuses on key issues of; openness, accountability, explanation, and responsibility. As seen in the literature review of this thesis, the link between bias and opacity means that the GDPR does to some degree address the issues of fairness and bias through tackling the question of opacity. However, the key legislation and regulation for protected characteristics within the Equality Act 2010 is much weaker when it comes to the oversight of AI decision-making. In this sense, the GDPR protects the individual, not groups, who may experience disparate impact in terms of bias in algorithmic decision-making.

By taking a sectoral approach to the review, the CDEI addresses some of the key issues raised within the literature review, namely that it is difficult to decouple the inherent bias which exists within institutions such as education with that of algorithmic decision making.

The CDEI takes the approach that, for example, when reviewing algorithmic-decision bias within recruitment, such a review must be part of the wider picture of how discrimination in recruitment is tackled more broadly (Centre for Data Ethics and Innovation 2020). In this sense, the work being conducted around bias adds value to a need to address equality and fairness across all aspects of UK society. Therefore, data and an ethical focus on it can be viewed as a tool or catalyst to create a better, fairer society with increased civic trust in public institutions.

Sectors covered in the review include; recruitment, financial services, policing and, local government. A common challenge and theme found across the review was that of governance. How is the risk of bias being anticipated and managed? Concerning governance, the report highlights the need for anticipatory governance. As aligned with the literature review, when considering The Singularity, good governance in this space must address a range of potential future impacts and intervene before issues occur. In this sense, anticipatory governance, as opposed to top-down policymaking, is recommended as providing a solution.

The review states that to improve fairness in decision making, it is critical that organisations need to be able to identify and address it. With this in mind, the review recommends:

- A need for diversity in the workforce
- Access to the right data is needed.
- Access to tools and approaches to identify and mitigate bias
- An ecosystem of experts is required to support organisations.
- Governance structures that anticipate risks and consider wider impact need to be in place
- Support of leadership and regulatory bodies to build confidence that a lawful and ethical approach is being taken.

In terms of creating a diverse workforce, the CDEI states that continued investment by the UK government in relevant skills development is required. For industry, the CDEI emphasises the need for organisations to make the diversity of teams and their workforce a priority.

Collecting the correct data is addressed in the review—recommendations are made relating to a need for a better understanding of data protection and collecting and sharing protected

characteristics. The review includes details as to why the collection of protected characteristics is needed to monitor and evaluate decision-making bias, by way of creating a baseline measure. Recommendations are also made around the sharing of service user data with trusted bodies.

The issue around fairness and the "mathematisation of ethics" is covered in detail in the review. Included is a recommendation to organisations in the UK using tools developed to statistically address fairness in the U.S. to be mindful that the equality law between the two sovereign states differs. Bias mitigation tools are in their early stages, and the question of legality across borders remains. With this in mind, the report concludes that concerning issues of fairness, further research is required and that holistic thinking is needed. The decision as to what is fair can not be left to data scientists to decide alone. It is confirmed that technical guidance for responsible bias detection and mitigation is needed, alongside cross-cutting guidance on the Equalities Act.

For the regulatory environment, the review highlights the gaps in EHRC capabilities. There is a need to build such capabilities to provide better guidance with direct relation to use of data and AI. Furthermore, the review highlights the need to include algorithmic bias as part of responsibilities under the PSED. In essence, the report calls for the adaptation and evolution of existing regulatory bodies and processes instead of creating a specific algorithmic regulatory system.

The issue of transparency and what should vs should not be published is raised. This value of openness is stated to be covered in the PBG as laid by in the previous section. This framework from the Alan Turing Institute is highlighted as a good starting point for all looking to implement use of data and AI ethically. Regarding explainability, the review states that the algorithmic element of decision-making should not be unexplainable and untransparent to the point that the public sector organisation using the tool cannot provide information about the whole decision process. Explainability, therefore, essentially rules out the use of the "black-box" decision-making approach as raised in the literature review

to be currently possible in achieving. Better solutions around explainability and use of black-box decision-making need to be found.

In summary, the challenges are moving at a fast pace. There are gaps in regulation, specifically around the incorporation of and interpretation of the Equality Act. There is also a landscape and good starting ground of existing regulation, legislation and guidance. Legislation, regulation and existing bodies can adapt with good leadership. The review concludes that the steps required to address bias in algorithmic decision-making align and overlap with those steps for tackling other ethical challenges, namely, good governance, data sharing, and explainability.

The CDEI review concludes by stating the need for an ecosystem of skilled professionals and supporting services to help organisations get fairness right.

4.7 Summary

In reviewing the current picture of data, AI and ethics in practice concerning the UK and public sector use of it, it is clear that there is great alignment with the issues being debated in the literature and the issues being tackled "on the ground."

One key aspect that stands out is the public sector's duty to proactively address inequality instead of maintaining the current "state of play." New technologies introduced must aid and not hinder the mission to deliver the Nolan Principles and the PSED. The issue of fairness; how to quantify and agree on definitions remains unsolved. Fairness combined with gaps around processes for addressing bias concerning protected characteristics as laid down within the Equality Act 2010 is the most pressing current gap and barrier to adopting new technologies.

The need for constant adaptation and evolution of existing institutions is highlighted as a key solution. In this sense, the gaps and barriers of today will not be those of tomorrow. In

order to maintain the trust of civil society in the long term, it is apparent that this evolution and adaptation will have to be constant. Suppose the trust of civil society is lost. In that case, the area of data, AI and ethics becomes not simply a case of addressing fairness and equality, but indeed key in preserving and maintaining democracy. It is, in this sense, a democratic issue, with a need for continued policy and good governance responses.

The following section will focus on potential solutions to the challenges faced by governments when considering policy and good governance responses concerning AI and algorithmic decision making in relation to bias.

Specifically, the next section will review the gaps, and thus barriers, highlighted by the review of public sector use of data and AI within the UK. Namely, the issue surrounding protected characteristics, the concept of fairness, the need to maintain the trust of civil society and the requirement for a measurable futures and outcome-focused approach to risk mitigation in relation to bias.

Chapter 5

Solutions to Algorithmic Bias - The Nonprofits

“There can be no algorithmic accountability without a critical audience. By this, I mean that, unless it draws the attention of people who critically engage with it, technical and non-technical quality assurance of algorithms is a token gesture and will fail to have the desired effect” (Kolkman, 2020)

When conducting the review of AI and Ethics concerning the UK public sector, the nonprofit sector plays a key and usual role in providing checks and balances. In this sense, organisations such as Liberty and Big Brother Watch are highlighted as rising awareness of risk through challenging practices. Nonprofits' profile is mostly noted when considering the issue of bias, where their presence appears strongest. Nonprofits are increasingly using the law courts to advance the rights of the groups and individuals they exist to serve in terms of addressing social justice. The gap in guidance for issues relating to bias as raised by the CDEI could further hamper progress in the use of AI. Key public services may wish to sit back and wait for the legal framework to become clear through watching cases play out in the courts.

As stated in the review conducted by the CDEI, the lens of the situation is that it is difficult to decouple the inherent bias which exists in society with that of algorithmic decision making. With this lens, it is then possible to widen the catchment of potential nonprofit engagement. In essence, those nonprofit actors addressing inherent bias within society for those with protected characteristics can indeed become viewed as potential partners in the solution, as opposed to simply the nonprofits providing the checks and balances with expertise in the area of data, AI, and algorithmic decision making. The issue of bias in algorithmic decision-making then becomes not just the domain of nonprofits operating within the space of social justice and human rights alone.

5.1 The rise of the Theory of Change

Since the 1990s, the nonprofit sector has witnessed a great change concerning how it conducts social innovation. This shift occurred because, as issues being tackled by the nonprofit sector became more complex, the ability to monitor and evaluate change and impact became increasingly difficult. This situation is a critical failure for the nonprofit business model, which is measured to be effective not by sales of a product or service but by the societal impact it can make.

In 1995 the Aspen Institute and its Roundtable on Community Change published *New Approaches to Evaluating Comprehensive Community Initiatives*. This publication stated that the key reason for the inability to conduct a robust evaluation on complex social impact programmes is the lack of detail surrounding the change process itself. The book highlighted the need to measure short and medium-term changes in addition to long-term goals. At this time, the nonprofit sector was more often than not using logframes to develop and evaluate social interventions. Logframes do not detail the short and medium-term outcomes, the change process required when assessing if, how, when, and why social impact has been made. The inability to assess the change process prevents effective scaling of social solutions.

Within this publication, *New Approaches to Evaluating Comprehensive Community Initiatives*, the term "theory of change" was first popularised. The Theory of Change (ToC) uses backward mapping, taking a long-term impact goal and developing a series of short, medium, and long-term outcomes required for the goal to be realised. Causal loops and pathways are also mapped. Actions and activities required to realise each series of outcomes are plotted into the ToC. Evaluation methods for each outcome are then devised. This assists with tracking the change caused by a whole project or intervention and measuring whether expected outcomes are actually realised (Weiss et al. 1995).

Since the publication of the book *New Approaches to Evaluating Comprehensive Community Initiatives*, the area of work around the development of the ToC has evolved.

The ToC is now linked to systems thinking. Change is no longer visualised as a linear process. Today, many toolkits, guides, and training programmes exist for the nonprofit sector on how to develop a ToC. The UK think tank New Philanthropy Capital has developed numerous toolkits for the ToC. Toolkits are applicable for those operating in system change, campaigning, project interventions, and grant-giving. The approach has grown so exponentially there is now even a nonprofit dedicated to delivering the process called the Centre for Theory of Change. An Example of completed two ToC's can be found in annex 2, with the accompanying information found in the full ToC report located in the bibliography.

5.2 Key Building Blocks of the Theory of Change

During creation, the ToC captures assumptions about what needs to be in place for a change to occur. The ToC also captures contextual factors, characteristics unique to a particular group, society, community, or individual. The inclusion of context characteristics and assumptions enables the nonprofit to challenge deeper beliefs, values, and operational 'rule of thumb' (Vogel 2012). The ability to capture and evaluate assumptions alongside multi-stakeholder engagement at the development stage helps challenge viewpoints, power relations, political, social, and environmental realities. Making assumptions explicit enables the constant monitoring of them throughout the delivery stages of the change process.

This element of the ToC process enables increased engagement with beneficiaries or the target group for whom the change is required. The National Council for Voluntary Organisations (NCVO), the umbrella body for the voluntary and community sector in England, lists the need for four key criteria to be met when conducting a ToC. Embedded within this criteria is stakeholder engagement. Stakeholder engagement is key to ensuring the ToC is both credible and supported. The criteria listed by the NCVO are:

Credible – based on previous experience and insight from your different stakeholders or relevant research where appropriate

Achievable – you have the necessary resources to carry out the intervention

Supported – stakeholders will be involved in defining and agreeing the theory of change, which builds support for it

Testable – a complete but not over-complicated description of your work and its outcomes, with prioritised outcomes for measurement and indicators to collect data against them (Brennan 2020, ncvo).

It should be noted that stakeholders are an ecosystem themselves, and thus the term does not solely relate to beneficiary engagement. However, today, beneficiary engagement at all levels of operation and intervention conception and delivery is considered best practice within the UK nonprofit sector. This move to a more "bottom-up" approach within the sector can be evidenced by the numerous toolkits and advice available for the sector when planning beneficiary participation. For example, the extensive guidance offered by NCVO and large national funding bodies such as the National Lottery Community Fund who state the need to evidence community leadership and engagement throughout as a criterion to access funding.

The ability to evaluate change is key to the ToC model. Every outcome listed has evaluation indicators attached, used to operationalise the change wanted. Indicators can be complex and tend to focus on the change experienced by targeted individuals, groups, or communities; in this sense, it is a qualitative and quantitative evaluation which is gathered. Different evaluation methods are chosen per indicator. The ToC process is underpinned by data at the development stage and evaluation data throughout all stages of the change process.

A ToC requires an evidence base for proof of concept as a form of risk management. For example, a nonprofit looking to deliver community cohesion will likely need to create an environment of shared values. Various forms of deliberative democracy are evidenced as decreasing partisanship and increasing shared consensus chances. The nonprofit, in this

instance, may decide to embed within their intervention the five key criteria highlighted by James Fishkin required to deliver deliberative democracy by way of a "proof of concept."

Furthermore, to evidence community cohesion, the nonprofit may wish to measure levels of shared responsibility and how empowered target groups of citizens feel to act to improve their local community. To harness such collective power, the intervention in this instance would likely consider aspects such as Sherry Arnstein's Ladder⁶ of Citizen Participation. Consideration of the Ladder of Citizen Participation would help capture and ensure a high degree of citizen power is embedded within the change process to be delivered. Furthermore, in this hypothetical example, various existing design principles for capturing and using collective intelligence may also be embedded and used as proof of concept.

5.3 Theory of Change as a Solution to Algorithmic Bias

The ToC process can be considered a more advanced and detailed version of the equality impact assessment (EIA) used to monitor and deliver the PSED in the public sector.

Use of EIA's is not mandatory within the public sector but is advocated by Trade Unions. Those EIA's which are used tend to be more linear and akin to a risk register. An example template of an EIA is shown in annex 3.

If delivery of the PSED and EIA's are considered a solution by CDEI to solving bias in algorithmic decision making, then a ToC being delivered by a nonprofit, targeted specifically at groups with protected characteristics, can, in theory, be considered as a more in-depth solution.

The CDEI highlights, in the review summarised within this thesis, a need for Anticipatory Governance. In the literature review, the challenges linked to an anticipatory approach are

⁶ Sherry R. Arnstein's "A Ladder of Citizen Participation," Journal of the American Planning Association, Vol. 35, No. 4, July 1969, pp. 216-224.

discussed. The issue of predictive policing and the pitfalls which can create bias and unfairness are highlighted and demonstrated.

Remaining with policing, consider predictive methods of resource allocation within policing. If used alongside the delivery of a ToC by nonprofits, the risks could be alleviated. If it is assumed that the ToC chosen is focused upon making positive change for a group with protected characteristics, then the ToC can assist with highlighting when bias occurs in terms of outcomes within the change journey for the target group and can thus be viewed as a complementary governance process. Essentially, the ability of the ToC to capture, test, and measure assumptions in the change journey could go some way to assisting with the ability to highlight where, when, and how bias starts to occur as experienced by the target group.

This process of collecting better, qualitative data can then be utilised to challenge human bias that occurs in the development and implementation of algorithmic models, highlighted as a key issue within the literature review. This ability to take a more holistic approach, and access a wide range of data from target groups who experience disparate bias, can also address the lack of workforce diversity, highlighted by the CDEI as an immediate and ongoing issue. Workforce diversity by its nature takes a long time to address. The final result would be an improved approach to delivering equity within public sector digital transformation, much more in line with the PSED and Equality Act 2010.

The inclusion of the nonprofit sector, and thus citizens in public sector digital transformation in this way, links to the question raised in the literature review of 'Who knows? Who decides? Who decides who decides?' (Zuboff 2020). Combining the ToC approach with public sector transformation suddenly increases meaningful citizens' engagement, specifically, citizen groups who are more subject to bias and disparate impact.

Finally, the ability of public services to work with the nonprofit sector and through this partnership, essentially with citizens themselves, will assist in creating a more transparent

approach to digital transformation. Civil society will become more informed and empowered within the process, aiding with the maintenance of trust. In a ToC, it would be perfectly acceptable to make this assumption and therefore measure it to be true or not.

5.4 Summary

It is the traditional role within a democratic system of the nonprofit sector to provide checks and balances for governments. Working in partnership with nonprofits in new ways, through an outcomes and futures-centred approach will benefit the ability of the public sector to address gaps in process around bias. Such a partnership, through the ToC process will ensure such checks and balances are timely, evidenced, relevant, and citizen-powered. In this sense the ToC process provides a form of anticipatory governance as referenced by the CDEI as a solution to algorithmic decision-bias. With the implementation of processes such as the ToC in the nonprofit sector, a better, improved, partnership approach can certainly be explored.

Chapter 6

Conclusion

It is clear that a cycle of investment, talent development, entrepreneurship and interest in the AI industry has increased its use and progress. This accelerated cycle of innovation within the AI industry has lowered costs in terms of development and lowered timescales for deployment and implementation. In this sense, the technical revolution being experienced today can be considered to be one of the fastest.

As with previous technology revolutions, the most famous being the Industrial Revolution, this current shift being caused by new technologies, which includes AI will have a wide-reaching impact on society and commerce. Productivity gains can be realised. Business models are and will continue to be disrupted. Changing business models result in a need for new skills and competencies in the labour market.

There are also positive signs that access to modern technological infrastructure, and new technologies are successfully bringing down barriers. For example, the opening up of the opportunities presented by globalisation to the SME and micro business sector. When considering the structural make-up of inequality, such decentralisation and the advantages modern technological infrastructure and technologies bring to the SME and micro business sector, this certainly presents a positive picture for the diffusion of influence and opening up of opportunity wider than for a concentrated few.

This research has focused on AI and machine learning. AI and machine learning are, as of yet not, as widely accessible for those with limited investment and resources. Systems are centralised, and models become outdated if they are not trained regularly with new data sets. However, organisations, such as Microsoft, are trying to make AI decentralised and collaborative using blockchain. This will result in people easily running cost-effective

machine learning models on accessible devices such as laptops while collaboratively contributing to data and thus improved models.

This research has demonstrated that AI and machine learning does not come without risks. Jobs will be displaced, and bias AI systems will simply increase inequality. High tech surveillance, enabled by AI will impact upon the human rights of individuals and democratic functions such as activism and journalism. A recent piece of investigative journalism by the UK newspaper the Guardian investigated and produced a series of reports on the extent of state-sponsored surveillance using software called Pegasus, sold by the company NSO. Surveillance using this software has included everyone from individuals to World leaders. This has increased tensions and potential conflicts between nations (Kirchgaessner et al., 2021).

Placing the lens of focus onto the UK and the use of AI and algorithmic decision making within public services, it is possible to use this case study to see the effect current legislation and governance has on addressing potential and future risks highlighted in the literature.

For the UK, the GDPR, the Equality Act, and the Public Sector Equality Duty serve as strong central regulations in their own right when considering many of the ethical issues raised by AI and machine learning. These same key issues, such as openness, transparency, privacy, bias and accountability, are also underpinned in the corporate governance structure of the UK Public Sector, namely the Nolan Principles. Where the regulatory and legal frameworks are vague, such as the right for explanation with the GDPR, administrative law and the right to an appeal in UK law can bolster this, and reviews and guidance are currently underway to clarify this area.

The UK case study demonstrated that clear ethical guidance, in line with existing regulation, is critical when considering ethical use of data and AI. While excellent work has been done by the key UK bodies responsible for ensuring any use of AI, automation or

algorithmic decisions is ethical, for example, the Turing Institute and the CDEI, this guidance is vast. The guidance is also very technical in its nature and complex to understand. Furthermore,, several competing sets of guidance exist in the UK. The impact is that many public institutions, such as local authorities, are hesitant to implement AI and algorithmic decision making, but wait and watch issues and cases play out in the law courts as a different way of guidance. This can hinder progress and the benefits the technology offers being released.

The issue of bias has not been resolved by the current UK guidance. The PSED and the Equality Act present a good regulatory framework in terms of preventing discriminatory practices for those with protected characteristics. The issue remains,, however, how do public bodies detect discrimination, what types of algorithmic profiling are discriminatory and when is bias to mitigated, and when should it remain? Whether these questions can be solved by better interpretation and evolution of the Equality Act remains unknown.

The final section of the research assessed the potential development of partnership working with the nonprofit sector, as a solution to the yet unanswered questions around guidance in relation to bias and discrimination. The public sector has a legal duty to proactively address issues of inequality. It remains a sad anomaly that inequality and prejudice surrounding certain groups remain within UK society. Through better working partnerships with the nonprofit sector, enabling the ToC process to compliment the use of EIAs, a more holistic approach to detecting discrimination and bias in algorithmic decisions could be sourced.

The ability to use the nonprofit sector outcomes approach and constant monitoring of the change journey process will serve to highlight when and how discrimination in the algorithmic process actually happens for target groups. This will assist the public sector in better highlighting and implementing mitigation techniques. Such an approach will also be led by those for whom the discrimination is actually happening to. Diversity of the workforce is not renowned in the AI industry and remains a highlighted issue. Increased and better quantitative evaluation and case studies gathered through the nonprofit ToC

process will also add value to a better, more diverse and holistic thinking within the AI industry and profession. This can, in turn, assist with challenging human bias, which impacts the technology at the design and implementation stage. As the literature review concluded, the technology itself is values neutral, but it is a human bias that can impact it.

Finally, the issue of fairness and how to build this into mathematical systems remains. What is fair to the author of this research may not feel fair to the reader of it. Fairness is subjective. However, collective ideas of fairness are built upon values. The public service in the UK is built upon the values of the Nolan Principles. With ethical guidance and governance that embodies the Nolan Principles and enables them to be continually realised, then algorithmic decision making in the UK should remain ethical and fair. Another success of the Nolan Principles in public service is that they are known, embedded and understood. The guidance provided by the key institutions for data, ethics and AI could perhaps follow this and promote access to and understanding of their guidelines more widely.

The UK witnessed an *annus horribilis* in 2020 in terms of high profile algorithmic disasters. However, there has been time, effort, thought, and investment in the area of data, ethics and AI by the UK Government. This level of commitment must continue in such a fast-moving space. The UK will never match the likes of the US or China in terms of investment spend in this area, however in the space of data, AI, and ethics, the country could carve itself a leading global role, and its vast and varied nonprofit sector could be the partners it needs to help it do so.

Bibliography

- Acemoglu, D., 2021. Remaking the post-COVID world. Remaking Post-COVID World. URL
<https://www.imf.org/external/pubs/ft/fandd/2021/03/pdf/COVID-inequality-and-automation-acemoglu.pdf>
- AI & Robotics Researchers, 2015. Open Letter on Autonomous Weapons [WWW Document]. Future Life Inst. URL
<https://futureoflife.org/open-letter-autonomous-weapons/> (accessed 7.16.21).
- Amer, K., Noujaim, J., 2019. The Great Hack. Netflix.
- Amini, A., Soleimany, A.P., Schwarting, W., Bhatia, S.N., Rus, D., 2019. Uncovering and Mitigating Algorithmic Bias through Learned Latent Structure, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. Presented at the AIES '19: AAAI/ACM Conference on AI, Ethics, and Society, ACM, Honolulu HI USA, pp. 289–295. <https://doi.org/10.1145/3306618.3314243>
- Arnstein, S., 1969. A Ladder of Citizen Participation. *Journal Am. Plan. Assoc.* 35, 216–224.
- Asaro, P.M., 2019. AI Ethics in Predictive Policing: From Models of Threat to an Ethics of Care. *IEEE Technol. Soc. Mag.* 38, 40–53. <https://doi.org/10.1109/MTS.2019.2915154>
- Brennan, R., 2020. How to build a theory of change — NCVO Knowhow [WWW Document]. NCVO. URL
<https://knowhow.ncvo.org.uk/how-to/how-to-build-a-theory-of-change> (accessed 7.23.21).
- Biotech and Biological Sciences Research Council, n.d. Equality impact assessment guidance and template 5.
<https://bbsrc.ukri.org/documents/equality-impact-assessment-guidance-template-pdf/>
(accessed 7/17/21).
- Brundage, M., 2015. Taking superintelligence seriously. *Futures* 72, 32–35.
<https://doi.org/10.1016/j.futures.2015.07.009>

Buchanan, B.G., 2005. A (Very) Brief History of Artificial Intelligence. *AI Mag.* 26, 53–53. <https://doi.org/10.1609/aimag.v26i4.1848>

Central Digital and Data Office, Office for Artificial Intelligence, 2019. A guide to using artificial intelligence in the public sector [WWW Document]. GOV.UK. URL <https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector> (accessed 7.21.21).

Centre for Data Ethics and Innovation, 2020. Review into bias in algorithmic decision-making [WWW Document]. GOV.UK. URL <https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making/main-report-cdei-review-into-bias-in-algorithmic-decision-making> (accessed 7.21.21).

Chauhan, G., 2013. A Review of Privacy Techniques. *Int. J. Eng. Res. Technol.* 2.

Closer to the Truth, 2021. Christof Koch - Must the Universe Contain Consciousness?, Must the Universe Contain Consciousness.

Committee on Standards in Public Life, 1995. The Seven Principles of Public Life [WWW Document]. GOV.UK. URL <https://www.gov.uk/government/publications/the-7-principles-of-public-life/the-7-principles-of-public-life--2> (accessed 7.20.21).

Covell, J., 2019. Project Explain interim report 31.

Dastin, J., 2018. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters.

Data Protection Act, 2018. Data Protection Act 2018 [WWW Document]. URL <https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted> (accessed 7.21.21).

DCMS, 2020. Data Ethics Framework [WWW Document]. GOV.UK. URL <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020> (accessed 7.21.21).

Department for Business, Energy and Industrial Strategy, 2021. The Grand Challenges [WWW Document]. GOV.UK. URL

<https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/industrial-strategy-the-grand-challenges> (accessed 7.20.21).

Diakopoulos, N., 2014. Algorithmic Accountability. *Digit. Journal.* 3, 398–415.
<https://doi.org/10.1080/21670811.2014.976411>

Duisberg, D.A., Vogel, D.B., 2017. Ethics Committee of German Federal Ministry of Transport and Infrastructure publishes guidance notes on automated driving [WWW Document]. *Bird Bird.* URL
<http://www.twobirds.com/en/news/articles/2017/germany/guidance-notes-on-automated-driving-published> (accessed 7.16.21).

Equality and Human Rights Commission, 2011. Public Sector Equality Duty | Equality and Human Rights Commission [WWW Document]. URL
<https://www.equalityhumanrights.com/en/advice-and-guidance/public-sector-equality-duty> (accessed 7.21.21).

European Commission. Joint Research Centre., 2017. What makes a fair society?: insights and evidence. Publications Office, LU.

Ewert, B., 2020. Moving beyond the obsession with nudging individual behaviour: Towards a broader understanding of Behavioural Public Policy. *Public Policy Adm.* 35, 337–360. <https://doi.org/10.1177/0952076719889090>

Feldman, N., 2018. Artificial Intelligence in Policing: Advice for New Orleans and Palantir - Bloomberg [WWW Document]. URL
<https://www.bloomberg.com/opinion/articles/2018-02-28/artificial-intelligence-in-policing-advice-for-new-orleans-and-palantir> (accessed 7.26.21).

Frankenstein, 2021. . Wikipedia.

Gerdon, S., Katz, E., LeGrand, E., Morrison, G., Torres Santeli, J., 2020. AI Procurement in a Box (Toolkit). *World Economic Forum.*

Hamilton, I.A., n.d. Why it's totally unsurprising that Amazon's recruitment AI was biased against women [WWW Document]. *Bus. Insid.* URL
<https://www.businessinsider.com/amazon-ai-biased-against-women-no-surprise-sandra-wachter-2018-10> (accessed 7.7.21).

Hao, K., 2019. Police across the US are training crime-predicting AIs on falsified data [WWW Document]. MIT Technol. Rev. URL <https://www.technologyreview.com/2019/02/13/137444/predictive-policing-algorithms-ai-crime-dirty-data/> (accessed 7.26.21).

Harari, Y.N., 2018. Homo Deus: a brief history of tomorrow.

Harris, J.D., Waggoner, B., 2019. Decentralized and Collaborative AI on Blockchain, in: 2019 IEEE International Conference on Blockchain (Blockchain). Presented at the 2019 IEEE International Conference on Blockchain (Blockchain), IEEE, Atlanta, GA, USA, pp. 368–375. <https://doi.org/10.1109/Blockchain.2019.00057>

Henley, P., 2021. The pandemic brings more robots. Real Story.

Institute for the Future, 2020. Anticipatory Governance.

Kahneman, D., 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics. *Am. Econ. Rev.* 93, 1449–1475. <https://doi.org/10.1257/000282803322655392>

Kirchgaessner, S., Lewis, P., Pegg, S., Lakhani, N., Cutler, S., Safi, M., 2021. Revealed: leak uncovers global abuse of cyber-surveillance weapon [WWW Document]. the Guardian. URL <http://www.theguardian.com/world/2021/jul/18/revealed-leak-uncovers-global-abuse-of-cyber-surveillance-weapon-nso-group-pegasus> (accessed 7.27.21).

Klein, C., 2021. The Original Luddites Raged Against the Machine of the Industrial Revolution [WWW Document]. HISTORY. URL <https://www.history.com/news/industrial-revolution-luddites-workers> (accessed 6.20.21).

Kolkman, D., 2020. “F**k the algorithm”?: What the world can learn from the UK’s A-level grading fiasco. *Impact Soc. Sci.* URL <https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/> (accessed 7.27.21).

Kurzweil, R., 2006. The singularity is near: when humans transcend biology. Penguin Books, New York, NY.

Lamb, E., n.d. Review: Weapons of Math Destruction [WWW Document]. *Sci. Am. Blog Netw.* URL

<https://blogs.scientificamerican.com/roots-of-unity/review-weapons-of-math-destruction/> (accessed 6.24.21).

Leslie, D., 2019. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. Zenodo. <https://doi.org/10.5281/ZENODO.3240529>

LeVine, S., 2017. Artificial intelligence pioneer says we need to start over [WWW Document]. Axios. URL <https://www.axios.com/artificial-intelligence-pioneer-says-we-need-to-start-over-1513305524-f619efbd-9db0-4947-a9b2-7a4c310a28fe.html> (accessed 6.24.21).

Magee, K., 2018. Top 10 quotes from our Cambridge Analytica whistleblower interview [WWW Document]. Campaign. URL https://www.campaignlive.co.uk/article/top-10-quotes-cambridge-analytica-whistleblower-interview/1497690?utm_source=website&utm_medium=social (accessed 6.22.21).

Mahoney, P., 2020. Schrems II: What Happened, Where Are We Now and Where Are We Going? • Apex Privacy. Schrems II - What Happened. URL <https://apexprivacy.com/what-happened-where-are-we-now-and-where-are-we-going/> (accessed 6.22.21).

Marcus, G., 2018. Deep Learning: A Critical Appraisal. ArXiv180100631 Cs Stat.

Marsh, S., 2020. Councils scrapping use of algorithms in benefit and welfare decisions [WWW Document]. the Guardian. URL <http://www.theguardian.com/society/2020/aug/24/councils-scrapping-algorithms-benefit-welfare-decisions-concerns-bias> (accessed 7.28.21).

Miller, K.K., Johnson, T.G., 2009. The Role of Agriculture and Farm Household Diversification in the Rural Economy of the United States. <https://doi.org/10.13140/RG.2.2.15334.83522>

Noble, J., De Ossorno Garcia, S., Gillman, A., 2021. How adults use the Kooth “positive virtual ecosystem.” <https://www.thinknpc.org/resource-hub/theory-of-change-kooth-for-adults/> (accessed 7/26/21)

OECD, 2021a. State of implementation of the OECD AI Principles: Insights from national AI policies (OECD Digital Economy Papers No. 311), OECD Digital Economy Papers. Paris. <https://doi.org/10.1787/1cd40c44-en>

OECD, 2021b. Government at a Glance 2021. OECD Publishing.

O’Neil, C., 2016. Weapons of math destruction: how big data increases inequality and threatens democracy, First edition. ed. Crown, New York.

Orlowski, J., 2020. The Social Dilemma. Netflix.

Phillips, P.J., Pohl, G., 2021. Algorithms, human decision-making and predictive policing. SN Soc. Sci. 1, 109. <https://doi.org/10.1007/s43545-021-00109-6>

Reedy, C., 2017. Kurzweil Claims That the Singularity Will Happen by 2045 [WWW Document]. Futurism. URL <https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045> (accessed 7.16.21).

Reuben, B., 2020. Guidance on AI and data protection [WWW Document]. Inf. Comm. Off. URL <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/> (accessed 7.21.21).

Reynoso, R., 2021. A Complete History of Artificial Intelligence [WWW Document]. G2. URL <https://www.g2.com/articles/history-of-artificial-intelligence> (accessed 6.18.21).

Schneier, B., 2016. Data and Goliath: the hidden battles to collect your data and control your world, First published as a Norton paperback. ed. W.W. Norton & Company, New York London.

Searle, J.R., 1980. Minds, brains, and programs. Behav. Brain Sci. 3, 417–424. <https://doi.org/10.1017/S0140525X00005756>

Smith, S., 2017. Eight top tips for beneficiary involvement — NCVO Knowhow [WWW Document]. URL <https://knowhow.ncvo.org.uk/organisation/strategy/beneficiaries/eighttips> (accessed 7.23.21).

Soper, S., 2021. Fired by Bot at Amazon: ‘It’s You Against the Machine.’ Bloomberg.com.
Sparrow, R., 2007. Killer Robots. *J. Appl. Philos.* 24, 62–77.
<https://doi.org/10.1111/j.1468-5930.2007.00346.x>

Summers, L.H., 2016. *The Age of Secular Stagnation*.

The European Parliament and The Council Of The European Union, 2016. 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). *Off. J. Eur. Union*.

Traité des sensations, 2020. . Wikipédia.

USDA - National Agricultural Statistics Service - About NASS - History of Agricultural Statistics [WWW Document], 2018. . *Natl. Agric. Stat. Serv.* URL https://www.nass.usda.gov/About_NASS/History_of_Ag_Statistics/index.php (accessed 7.10.21).

Vogel, I., 2012. Review of the use of ‘Theory of Change’ in international development 86.

Weiss, C., Kubisch, A.C., Connell, J.P., Schorr, L. (Eds.), 1995. *New approaches to evaluating community initiatives*. Aspen Institute ; [Order from] Aspen Institute, Publications Office, Washington, D.C. : Queenstown, MD.

What is IoT? | IoT & Sensor Technology | OMRON - Americas - Americas [WWW Document], n.d. URL <https://components.omron.com/sensor/about-iot> (accessed 6.21.21).

Whitcomb, C.G., 2020. Review of Shoshana Zuboff (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*: New York: Public Affairs. 704 pp. ISBN 9781781256848 (Hardcover). *Postdigital Sci. Educ.* 2, 484–488.
<https://doi.org/10.1007/s42438-019-00086-3>

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., Myers West, S., Richardson, R., Schultz, J., Schwartz, O., 2018. *AI Now 2018 Report*.

Williams, E., 2020. THE UK: A PARABLE OF DISTRUST [WWW Document]. Edelman. URL <https://www.edelman.com/insights/uk-parable-distrust> (accessed 7.28.21).

Zuboff, S., 2020. The age of surveillance capitalism: the fight for a human future at the new frontier of power, First Trade Paperback Edition. ed. PublicAffairs, New York.

Annex 1- Infographic for a history of AI


Artificial Intelligence: A Brief History

1950	<p>Alan Turing publishes "Computing Machinery and Intelligence". John McCarthy develops Lisp. Arthur Samuel develops checkers playing computer programme and creates phrase "machine learning". Allen Newell, Herbert Simon, and Cliff Shaw co-author Logic Theorist.</p>	
1960	<p>General Motors use a robot on their assembly line James Slagle develops SAINT (Symbolic Automatic INTeegrator). Daniel Bobrow creates STUDENT Shakey the Robot developed by Charles Rosen. Terry Winograd creates SHRDLU. Edward Feigenbaum and Julian Feldman publish "Computers and Thought".</p>	
1970	<p>WABOT-1 built in Japan British government reduce funding for AI due to James Lighthill's report The Stanford Cart becomes the first autonomous vehicle. George Lucas releases Star Wars.</p>	
1980	<p>WABOT-2 built Japan allocate \$850 million to the Fifth Generation Computer project. AI Winter of low funding and interest in AI. Mercedes-Benz build and release a driverless van equipped. Judea Pearl publishes "Probabilistic Reasoning in Intelligent Systems." Chatbot Jabberwacky communicates with people</p>	

Annex 1 Continued


1990

Richard Wallace develops chatbot A.L.I.C.E.
Sepp Hochreiter and Jürgen Schmidhuber develop Long Short-Term Memory (LSTM).
IBM's Deep Blue wins at chess
Furby, first robot toy launched
Sony introduce AIBO, a robotic pet dog.




2000 to 2010

Y2K (the year of 2000 problem)
Cynthia Breazeal develops Kismet.
Honda releases ASIMO, an AI humanoid robot.
i-Robot releases Roomba
Opportunity and Spirit navigate Mars for NASA
Oren Etzioni, Michele Banko, and Michael Cafarella coin the phrase "machine reading"
Fei Fei Li and colleagues assemble ImageNet
Google secretly develop a driverless car
The films A.I. Artificial Intelligence and I,Robot are released



2010 to now

ImageNet launched the ImageNet Large Scale Visual Recognition Challenge
Microsoft launched Kinect for Xbox 360
Watson created by IBM
Apple release Siri
Jeff Dean and Andrew Ng train a large neural network of 16,000 processors to recognize images of cats
Carnegie Mellon University release Never Ending Image Learner (NEIL).
Microsoft release Cortana.
Amazon release Alexa.
Elon Musk, Stephen Hawking, Steve Wozniak and 3,000 others call for a ban to use of autonomous weapons
A humanoid robot named Sophia is created by Hanson Robotics.
Google release Google Home
The Facebook Artificial Intelligence Research Lab train two "dialog agents" (chatbots) to communicate with each other.
Alibaba language processing AI outscores human intellect at a Stanford
Google developed BERT
Samsung launch Bixby

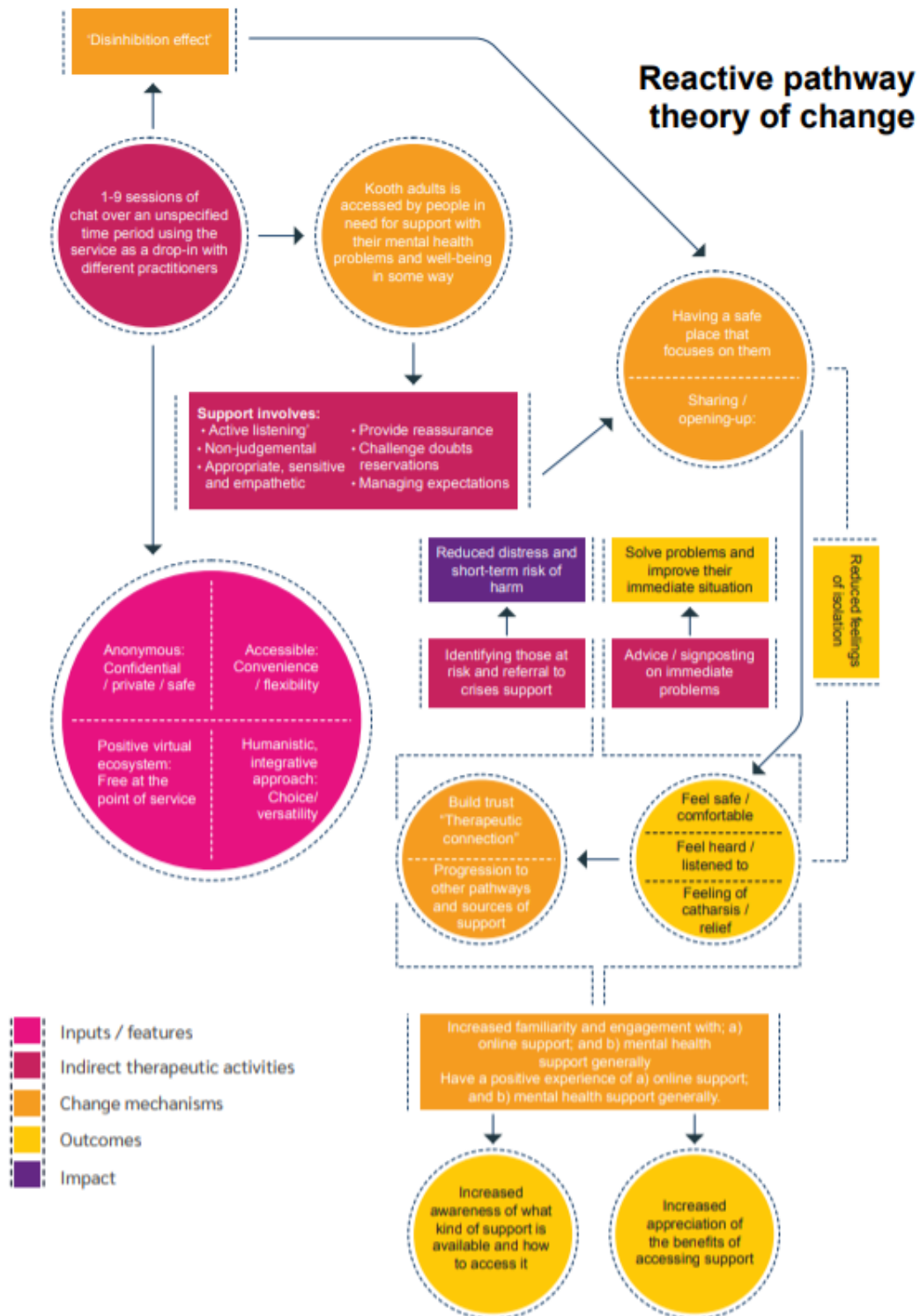


Sources: Buchanan, Bruce G. 2005. "A (Very) Brief History of Artificial Intelligence." AI Magazine 26 (4): 53–53. <https://doi.org/10.1609/aimag.v26i4.1848>.
Reynoso, Rebecca. 2021. "A Complete History of Artificial Intelligence." G2. May 25, 2021. <https://www.g2.com/articles/history-of-artificial-intelligence>.

Sources: Buchanan, Bruce G. 2005. "A (Very) Brief History of Artificial Intelligence." AI Magazine 26 (4): 53–53. <https://doi.org/10.1609/aimag.v26i4.1848>.

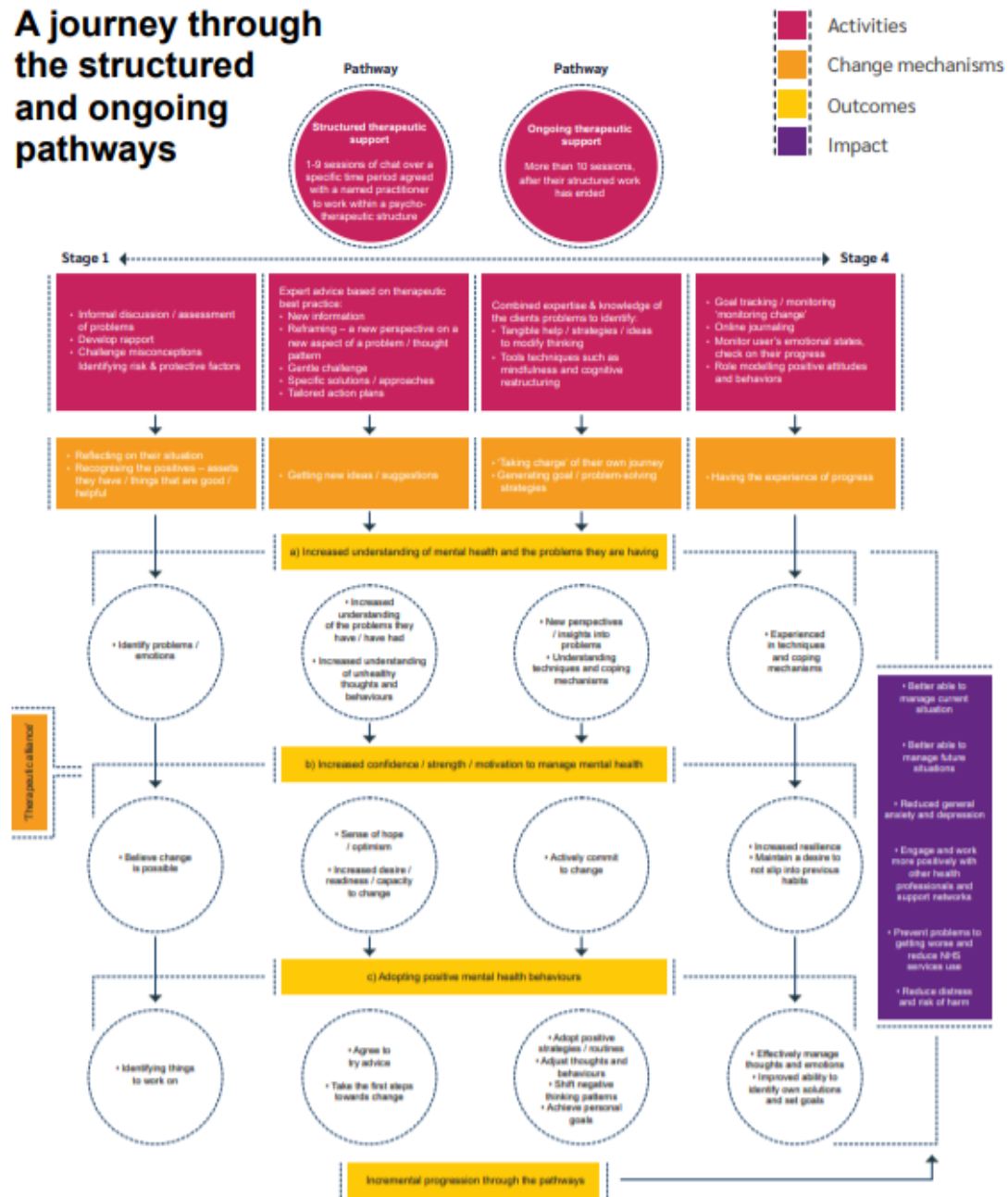
Reynoso, Rebecca. 2021. "A Complete History of Artificial Intelligence." G2. May 25, 2021. <https://www.g2.com/articles/history-of-artificial-intelligence>.

Annex 2 - Example of a nonprofit Theory of Change (ToC)



Annex 2 Continued

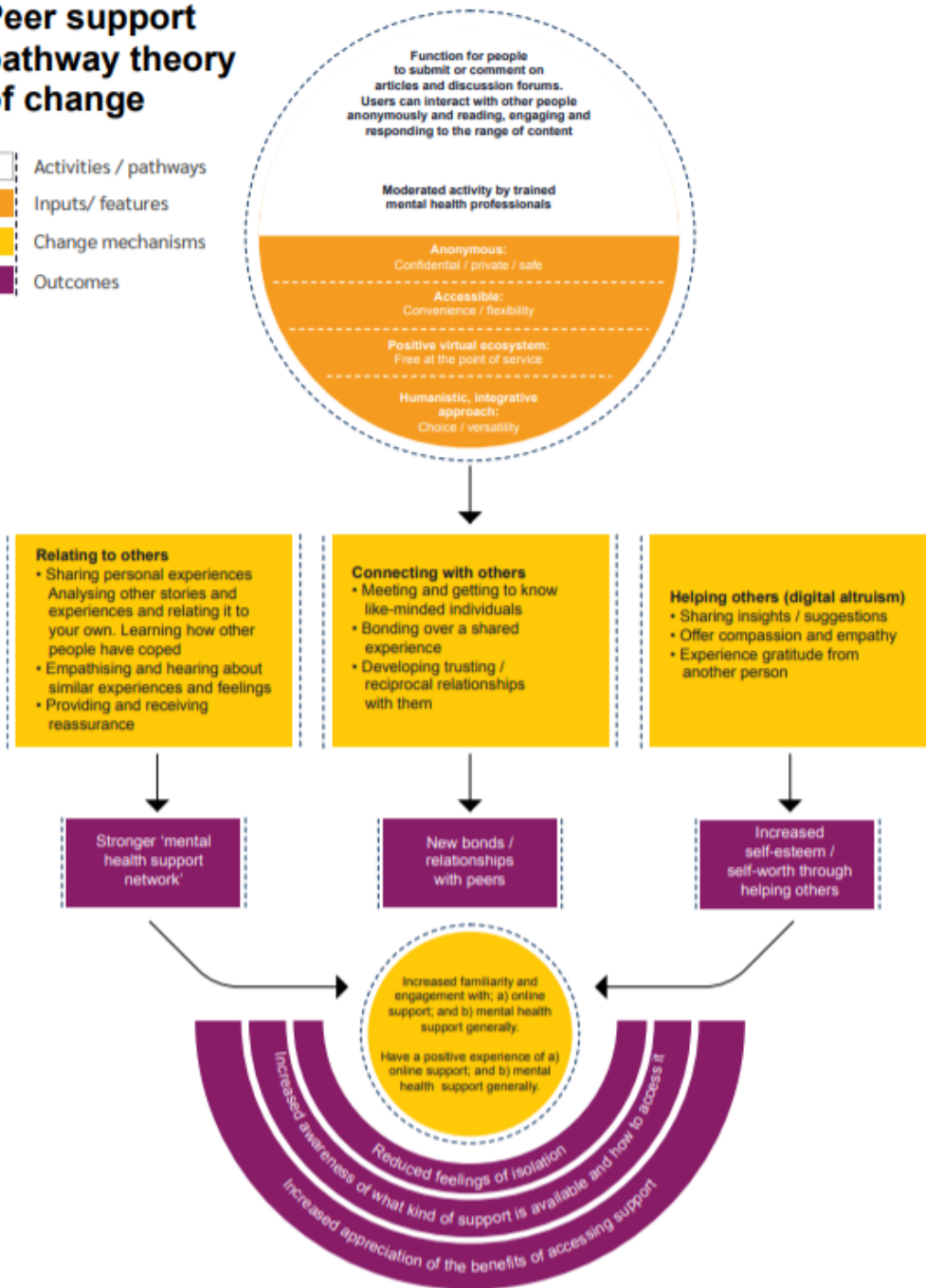
A journey through the structured and ongoing pathways



Annex 2 Continued

Peer support pathway theory of change

- Activities / pathways
- Inputs/ features
- Change mechanisms
- Outcomes



(Noble et al., 2021)

Annex 3 - Equality Impact Assessment (EIA)

Equality Impact Assessment

Question	Response
1. Name of policy/funding activity/event being assessed	
2. Summary of aims and objectives of the policy/funding activity/event	
3. What involvement and consultation has been done in relation to this policy? (e.g. with relevant groups and stakeholders)	
4. Who is affected by the policy/funding activity/event?	
5. What are the arrangements for monitoring and reviewing the actual impact of the policy/funding activity/event?	

Protected Characteristic Group	Is there a potential for positive or negative impact?	Please explain and give examples of any evidence/data used	Action to address negative impact (e.g. adjustment to the policy)
Disability			
Gender reassignment			
Marriage or civil partnership			
Pregnancy and maternity			
Race			
Religion or belief			
Sexual orientation			
Sex (gender)			
Age			

Annex 3 Continued

Evaluation:

Question	Explanation / justification	
Is it possible the proposed policy or activity or change in policy or activity could discriminate or unfairly disadvantage people?		
Final Decision:	Tick the relevant box	Include any explanation / justification required
1. No barriers identified, therefore activity will proceed .		
2. You can decide to stop the policy or practice at some point because the data shows bias towards one or more groups		
3. You can adapt or change the policy in a way which you think will eliminate the bias		
4. Barriers and impact identified, however having considered all available options carefully, there appear to be no other proportionate ways to achieve the aim of the policy or practice (e.g. in extreme cases or where positive action is taken). Therefore you are going to proceed with caution with this policy or practice knowing that it may favour some people less than others, providing justification for this decision.		

Annex 3 continued

Will this EIA be published* Yes/Not required (*EIA's should be published alongside relevant funding activities e.g. calls and events:	
Date completed:	
Review date (if applicable):	

Change log

Name	Date	Version	Change
	When published	1	

(Biotech and Biological Sciences Research Council, n.d.)