AN EQUINET PUBLICATION

2020

REGULATING FOR AN EQUAL AI: A NEW ROLE FOR EQUALITY BODIES

Meeting the new challenges to equality and non-discrimination from increased digitisation and the use of Artificial Intelligence

by Robin Allen QC and Dee Masters

EQUINET
European Network
of Equality Bodies

REGULATING FOR AN EQUAL AI: A NEW ROLE FOR EQUALITY BODIES

Meeting the new challenges to equality and non-discrimination from increased digitisation and the use of Artificial Intelligence

BY ROBIN ALLEN QC AND DEE MASTERS

Brussels | 2020

an Equinet publication

Co-funded by the Rights, Equality and Citizenship Programme of the European Union

Regulating for an Equal AI: A New Role for Equality Bodies - Meeting the new challenges to equality and non-discrimination from increased digitisation and the use of Artificial Intelligence is published by Equinet, European Network of Equality Bodies. Equinet brings together 49 organisations from across Europe which are empowered to counteract discrimination as national equality bodies across the range of grounds including age, disability, gender, race or ethnic origin, religion or belief, and sexual orientation. Equinet works to enable national equality bodies to achieve and exercise their full potential by sustaining and developing a network and a platform at European level.

Equinet members: Commissioner for the Protection from Discrimination, **Albania** | Austrian Disability Ombudsman, **Austria** | Ombud for Equal Treatment, **Austria** | Unia (Interfederal Centre for Equal Opportunities), **Belgium** | Institute for Equality between Women and Men, **Belgium** | Institution of Human Rights Ombudsman, **Bosnia and Herzegovina** | Commission for Protection against Discrimination, **Bulgaria** | Office of the Ombudsman, **Croatia** | Ombudsperson for Gender Equality, **Croatia** | Ombudswoman for Persons with Disabilities, **Croatia** | Office of the Commissioner for Administration and the Protection of Human Rights (Ombudsman), **Cyprus** | Office of the Public Defender of Rights, **Czech Republic** | Board of Equal Treatment, **Denmark** | Danish Institute for Human Rights, **Denmark** | Gender Equality and Equal Treatment Commissioner, **Estonia** | Ombudsman for Equality, **Finland** | Non-Discrimination Ombudsman, **Finland** | Defender of Rights, **France** | Public Defender (Ombudsman), **Georgia** | Federal Anti-Discrimination Agency, **Germany** | Greek Ombudsman, **Greece** | Equal Treatment Authority, **Hungary** | Office of the Commissioner for Fundamental Rights, **Hungary** | Irish Human Rights and Equality Commission, **Ireland** | National Office Against Racial Discrimination, **Italy** | Ombudsperson Institution, **Kosovo*** | Office of the Ombudsman, **Latvia** | Office of the Equal Opportunities Ombudsperson, **Lithuania** | Centre for Equal Treatment, **Luxembourg** | National Commission for the Promotion of Equality, **Malta** | Commission for the Rights of Persons with Disability, **Malta** | Council on Preventing and Eliminating Discrimination and Ensuring Equality, **Moldova** | The Protector of Human Rights and Freedoms (Ombudsman), **Montenegro** | Netherlands Institute for Human Rights, **Netherlands** | Commission for Prevention and Protection against Discrimination, **North Macedonia** | Equality and Anti-Discrimination Ombud, **Norway** | Commissioner for Human Rights, **Poland** | Commission for Citizenship and Gender Equality, **Portugal** | Commission for Equality in Labour and Employment, **Portugal** | High Commission for Migration, **Portugal** | National Council for Combating Discrimination, **Romania** | Commissioner for Protection of Equality, **Serbia** | National Centre for Human Rights, **Slovakia** | Advocate of the Principle of Equality, **Slovenia** | Council for the Elimination of Ethnic or Racial Discrimination, **Spain** | Institute of Women and for Equal Opportunities, **Spain** | Equality Ombudsman, **Sweden** | Equality and Human Rights Commission, **UK - Great Britain** | Equality Commission for Northern Ireland, **UK - Northern Ireland**

*This designation is without prejudice to positions on status, and is in line with UNSCR 1244/1999 and the ICJ Opinion on the Kosovo declaration of independence.

Equinet Secretariat | Rue Royale 138 | 1000 Brussels | Belgium | info@equineteurope.org | www.equineteurope.org

Equinet's main contact point for this publication: Milla Vidina, policy officer | Milla.Vidina@equineteurope.org
Proof-editing work: Sophie Hale; Layout and Pre-publication Preparations: Levente Kollár

978-92-95112-36-0 (Print) / 978-92-95112-35-3 (Online)

This publication was commissioned by Equinet and written by Robin Allen QC and Dee Masters (AI Law Consultancy/Cloisters). The views expressed in this publication belong to the author. Neither Equinet nor the European Commission are liable for use that may be made of the information contained therein. This information does not necessarily reflect the position or the opinion of the European Commission

EQUINET
European Network of Equality Bodies

Co-funded by the Rights, Equality and Citizenship Programme of the European Union

CONTENTS

PREFACE

It is my pleasure to introduce the report by the European Network of Equality Bodies (Equinet) on the role of equality bodies in ensuring that everyone in Europe can benefit from the increased use of Artificial Intelligence (AI) systems in a fair and non-discriminatory way.

Europe must lead the transition to a new digital world—this is one of the headline ambitions of the European Commission. AI, as the engine of this digital transformation, offers important efficiency and productivity gains that can strengthen the competitiveness of European industry and improve the wellbeing of citizens.

AI should bring people together and leverage all of our strengths, talent and potential. Therefore, **a European approach for AI must respect and promote equality for all and equality in all of its senses.** Only then, we can ensure that the future development of AI contributes to a prosperous and social Union for us all.

Technology is only as good as the humans developing it and AI is no exception. The same human biases, prejudices and stereotypes that lead to discrimination can be replicated in the data and codes used by AI. AI could even amplify these biases and create new categories of unjust exclusion. We must protect our societies and all individuals from these threats.

Therefore, it is imperative to place **equality and its effective protection through well-resourced and empowered equality bodies at the heart of any European approach to AI,** in line with the recommendations in this report. I am committed to continue the European Commission's work towards equality bodies being adequately and securely supported with the necessary mandate and resources, in accordance with the 2018 European Commission Recommendation on Standards for Equality Bodies. This is an essential precondition for ensuring that the development and use of AI in the EU is grounded in respect for equality and contributes to the wellbeing of all.

Helena Dalli

European Commissioner for Equality

May 2020

JOINT INTRODUCTION

In 2019, the Board of Equinet, the European Network of Equality Bodies, noted the lack of any European study on the benefits and risks to the principle of equality caused by automated decision making, and, more generally, by Artificial Intelligence (AI). Therefore, in its 2019 Work Programme, Equinet committed to commissioning a "study on the consequences of digitalisation for (in)equality and the role equality bodies can play in this field" with a view to filling this gap and triggering discussions across Europe, particularly among Equinet's Members, on the effect that AI-driven technologies are having, and will have, on the principle of equality.

The idea for the study originated before 2019 with the realisation within the Equinet Executive Board that the effective protection of equality in Europe by Equality Bodies is likely to be significantly affected by progressive digitalization through the use of AI systems. From the outset, it was anticipated that many Equality Bodies had a relatively limited understanding of the ways in which AI systems could impact equality and needed practical and actionable guidance as to how to apply and enforce existing equality legislation in situations involving the use of AI technologies. Thus, the study was envisioned as primarily an internal capacity building tool for Equinet's Members Equality Bodies across Europe. In the summer of 2019, Robin Allen QC and Dee Masters were commissioned to undertake this study and make a final presentation of their findings in early 2020.

This Report presents the resulting information. In accordance with the terms of the research assignment, the Report meets three key objectives: it maps and identifies 1) the equality implications of AI systems, 2) makes recommendations on the role of Equinet's Members in the public discourse on AI and algorithmic discrimination, and 3) provides practical guidance to these Members on identifying and assessing the equality-relevant consequences of AI and automated-decision making.

In the course of developing the Report and subsequent revisions, it became clear that in order to accomplish the above objectives, the study had to focus on the importance of partnerships between Equality Bodies and diverse actors from different sectors and jurisdictions at both national and European level. The very nature of AI-enabled technologies as general purpose technologies, which could have many different uses in many sectors as different as finance, data protection and product safety, means an exponential increase—both in number and in kind—of the potential sources of discrimination. To successfully respond to this new challenging context, Equality Bodies have to proactively

connect and forge stable partnerships with a number of actors, many of which might be considered non-traditional for the equality legal field, including, for example, sectoral regulators such as Data Protection Authorities and Consumer Protection Authorities, computer and data scientists and engineers, within both the private sector, academia, digital rights NGOs and standardization bodies.

Naturally, the significance of partnerships for the capacity of Equality Bodies to protect against AI-related risks meant modifying the original conceptualisation of the study as primarily an internal tool addressed to Equinet's own members. As a result, the present study also addresses external partners which it has identified as key to enabling Equality Bodies to provide effective protection against AI-specific threats to the principle of equality and non-discrimination.

This change underscores one of the substantive messages of this Report, which we consider central to its contribution: the task of protecting and promoting equality against threats of AI-enabled technologies is an intrinsically collaborative endeavour, which necessitates active partnership of national equality bodies with national governments, relevant public authorities, such as regulators, as well as actors working on the topic of equality and AI at the European level, such as Equinet, the European Union and the Council of Europe (CoE), specifically the CoE's Ad hoc Committee on Artificial Intelligence (CAHAI).

While Equality Bodies are best placed to take the lead in equality-specific oversight and enforcement in the context of AI technologies at a national level, the complex nature and cross-sectoral use of AI systems means that more than ever Equality Bodies have to rely on key partners such as, for example, national Data Protection Authorities and the scientific community spearheading AI developments.

Since this is such a fast-moving area of work, this Report must be seen as the product of this specific moment in the development of these AI technologies and their impact on equality. There is no doubt that this will change significantly during 2020, not least because it is expected that both the European Commission and the Council of Europe (particularly through its Ad hoc Committee on Artificial Intelligence – CAHAI) will move to regulate AI systems. These developments are highlighted in this Report. Where possible, the Report has included significant events occurring after the first presentation of the Report but before final publication.

Importantly, the Report notes the publication of the European Commission's White Paper on AI, with which the Commission launched an ongoing consultation for proposals for key elements of a future regulatory framework on AI. We welcome the ambition of the Commission set out in the White Paper to develop

a common European approach grounded in EU values and fundamental rights, including equality and non-discrimination. In this context, the timing of the Report's launch could not be more opportune as its conclusions and recommendations should be seen as responses to many of the key questions underlying the Commission's consultation. Indeed, this Report complements the White Paper, by noting highlights and remedies from the standpoint of the protection of equality and other fundamental rights, paying particular attention to the users' side and more specifically, the ability of users of AI systems, to identify and claim their rights through adequate enforcement and redress channels. In this, it sits in parallel to the the White Paper on AI, which primarily addresses the "supply side" such as AI developers, vendors, and distributors. It is understandable that there should be a focus on building in user protections, including in relation to equality and fundamental rights, on the supply side, nevertheless, it is precisely in the context of AI systems that this should be reinforced by focus on the enforcement of users' rights, including first and foremost their fundamental rights.

Firstly, this is because any future AI-related regulatory developments, which ensure AI's compliance with EU's fundamental rights and values, will depend on initially assessing the extent to which existing equality and fundamental rights legislation can be applied and enforced adequately to address the risks that AI systems can create. Thus, the content and scope of regulatory change will depend directly on the implementation and enforcement shortcomings of existing equality and fundamental rights legislation in the context of AI systems.

Secondly, effective enforcement and redress are the areas within human rights protection most negatively affected by AI-enabled technologies, thus making the role of independent enforcement and redress mechanisms such as Equality Bodies all the more essential. As highlighted by the White Paper on AI, this is so because the essential characteristics of many AI systems, such as opacity ('black box-effect'), complexity and unpredictability, make it harder for enforcement institutions and affected persons to verify whether a decision, made with the involvement of AI, was taken in compliance with equality and fundamental rights rules, thus also adversely impinging upon effective access to evidence and redress possibilities. Moreover, as the proliferation of AI systems in the EU (as well as in wider Europe and globally) swiftly gathers speed, the risks to persons who have suffered harm because of AI-related equality and fundamental rights violations are expected to increase exponentially, threatening to create a situation of widespread denial of access to justice.

Thus, by focusing on the ability of existing equality legislation to tackle the risks of AI and the capacity of Equality Bodies to ensure adequate application and enforcement of the law to address these risks, the present Report contributes

to an important change of perspective toward the rights of the users of AI technology. Equality Bodies have a leading role to play to secure the benefits of AI across Europe without causing any adverse effect on equality and human rights.

It is, therefore, our special privilege to be given this opportunity to introduce this important contribution on a topic that is already shaping our present and will inevitably shape the future of equality in Europe.

Tena Šimonović Einwalter

Chair of Equinet Executive Board

Robin Allen QC and Dee Masters

AI Law Consultancy, Cloisters

May 2020

ABBREVIATIONS

All abbreviations are noted in brackets in the main text at the point at which they are first used.

ADM	Automated decision making
AI	Artificial Intelligence. In this Report "AI" is used both specifically and generically. It will be clear from the context which meaning is intended. When used generically, as in for instance "AI systems", ML and ADM and other forms of computer algorithm derived outputs are also intended to be included.
AI HLEG	EU's High-Level Expert Group on Artificial Intelligence
CAHAI	Council of Europe Ad Hoc Committee on Artificial Intelligence
CDEI	Centre for Data Ethics and Innovation
CFREU	Charter of Fundamental Rights of the European Union
CJEU	Court of Justice of the European Union
CNIL	Commission nationale de l'informatique et des libertés
CoE	Council of Europe
DWP	Department for Work and Pensions
EC	European Commission
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
EDPB	European Data Protection Board
Equinet	Equinet Aisbl - European Network of Equality Bodies
EU	European Union
FRA	European Union Agency for Fundamental Rights Agency
FRT	Facial recognition technology
GDPR	General Data Protection Regulation: 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

IAPP	International Association of Privacy Professionals
IEEE	The Institute of Electrical and Electronics Engineers – also known as "I triple E"
LED	Law Enforcement Directive: Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA
MIT	Massachusetts Institute of Technology
ML	Machine learning
MSI-AUT	CoE Committee of experts on Human Rights Dimensions of automated data processing and different forms of artificial intelligence
NHRI	National Human Rights Institutions
NGO	Non-governmental organisation
PECD	Privacy and Electronic Communications Directive: Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector
RBV	Risk-Based Verification
TFEU	Treaty on the Functioning of the European Union
UN	United Nations
UNCRPD	UN Convention on the Rights of Persons with Disabilities

EXECUTIVE SUMMARY

Overview

This Report addresses the new challenge facing all European Equality Bodies of upholding the principles of equality and non-discrimination in their respective states and areas in the face of the rapid development of new technologies using Artificial Intelligence (AI), algorithms, machine learning (ML), and automated decision-making (ADM).[1]

The work towards writing this Report commenced with a survey designed by the authors in collaboration with the Equinet Secretariat. The results of this survey are set out in Appendix 3. A review of the survey was followed by further meetings with Equinet's Members, carried out over the latter part of 2019 which included meetings in Paris on 26 September 2019, Brussels on 23 October 2019 and Berlin on 25 November 2019. Supplementary research took place throughout the second half of 2019 and early 2020.

The key questions

In the process of preparing this Report it became clear that there were six key questions for Equinet and its Members. These questions have been considered at every stage in the preparation of this Report. The questions are –

- What tasks should Equinet's Members undertake to ensure that AI, ML and ADM advance and do not hinder equality and non-discrimination?
- What capacity do Equinet's Members have for this?
- How can they be assisted to gain better capacity?
- Who or what are the other actors in this field with which Equinet's Members should be working?
- Does the current discourse on the ethical approach to AI support legal rights to equality?
- How well do the other regulatory tools available in states work with equality rights in the context of AI?

The answers to these questions are complex and discussed in detail in the Chapters of this Report and its Appendices. They are not readily summarised in

1 "AI" is used both specifically and generically in this Report. It will be clear from the context which meaning is intended. When used generically, as in for instance "AI systems", ML and ADM and other forms of computer algorithm derived outputs are also included.

a few sentences. However, the approach to these questions will be outlined in the following summary of the Chapters and Appendices of this Report.

Chapter 1 sets the scene by explaining the nature of the new technologies, and the developing interest in addressing their effects by the European Union (EU) and the Council of Europe (CoE). It draws a distinction between Europe and the United States of America pointing out the different approaches to equality.

Chapter 2 explains the main common uses of AI systems that should be of specific concern to Equinet's Members. These concern employment, biometric identification through facial recognition technology (FRT), education, recruitment, predictive policing, immigration and border control, financial products, health, social advantages, child welfare, justice and criminal justice systems, fraud detection, and military systems. It also notes the speed with which change is happening and the potential for economic benefits that is thought to exist. Appendix 1 supplements this chapter, outlining the use of these technologies in respective states.

Chapter 3 considers the legal resources currently available to Equinet's Members in both an equality and data protection perspective. It explains how the existing equalities framework would be applied to four case studies: (1) concerning direct discrimination as a result of the use of a basic, relatively simple algorithm, (2) the use of FRT, (3) the prediction of risk, and (4) the determination of immigration status. The chapter considers both direct and indirect discrimination by the state and the evidential issues caused by the lack of transparency in AI systems, the so-called "black-box" problem. In discussing (2) the use of FRT, this chapter explains in detail the importance of understanding how ethical principles concerning AI will often determine the question whether there is indirect discrimination. This chapter also addresses the ways in which European data protection laws can be used to regulate discriminatory AI. There is also an analysis of initiatives within Equinet's Membership in both the data protection and discrimination field. The chapter concludes by arguing that Equinet's Members can play a pivotal role in further shaping the legal framework, both in relation to equality law and data protection rules.

Chapter 4 summarises the work that Equinet's Members are currently undertaking to meet the new challenges raised by AI systems. It corresponds closely with the results of the survey, set out in Appendix 3, but goes further, addressing the information obtained from discussions with Unia (Belgium), the Defender of the Rights (France), the Federal Anti-Discrimination Agency (Germany), the Human Rights and Equality Commission (Ireland), and the Institute for Human Rights (the Netherlands).

Chapter 5 outlines proposals for further action by Equinet's Members and their respective states. It advises Equinet's Members to (1) identify the resources they need in light of the unique challenges posed by AI, (2) ensure that their individual mandates are sufficiently broad to address the new challenges, (3) map the ways in which these new technologies can affect equality and non-discrimination principles within their states, and (4) identify any legal gaps that need to be addressed and in this regard the Report sets out the policy, legislative and regulatory issues that Equality Bodies should consider. It concludes that states must ensure that Equinet's Members are fully able to meet the challenges arising from AI, both in terms of their mandates, and the resources made available to them. It advises that states must address the legal issues that will arise from the work of Equinet's Members as summarised above.

Chapter 6 sets out a checklist for Equinet's Members to assist them in developing future work towards understanding the equality implications of AI systems.

Appendix 2 discusses important initiatives occurring across Europe in relation to the regulation of AI systems.

Ethical principles

In many places this Report discusses the ethical principles that are being developed in relation to the creation and use of AI. This is typical of how societies introduce new technologies. First the outline of technological change is identified; next engineers consider how this insight can be developed and used; thereafter questions about the ethical implications of these new technologies are raised; finally society decides how the technology should be regulated to conform to accepted ethical standards.

At present there is a world-wide discussion as to the ethical principles which are relevant to AI[2] leading into the next stage – a discussion about regulation. However, decisions as to regulation have not been finally made. This Report discusses that work in Appendix 2, noting both the CoE's April 2020 "Recommendation on the human rights impacts of algorithmic systems"[3] and the European Commission's February 2020 White Paper "On Artificial Intelligence – A European approach to excellence and trust".[4] Appendix 2 describes the steps being taken now by the European Commission (EC) to consider whether ethical

2 Algorithm Watch has produced a detailed "inventory" of current AI ethical principles. It is available here: https://inventory.algorithmwatch.org/ and demonstrates the sheer array of ideas in this area.

3 See https://search.coe.int/cm/pages/result_details.aspx?ObjectId=09000016809e1154

4 See https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

principles should be made into legal rules. This Report is not a full response to the EC's consultation on these issues, but it does identify many of the problems arising from the absence of specific laws regulating the use of AI in accordance with the principle of non-discrimination and equal treatment.

This Report also explains how commonly agreed ethical principles will interact with existing legal rules protecting the principle of non-discrimination. In particular, it shows how knowledge of such ethical principles is critical for determining whether there is unlawful indirect discrimination. This is explained in the discussion in Chapter 3 as to how Facial Recognition Technology can be unlawfully indirectly discriminatory, but it is emphasised that this example concerns only one instance of how ethical principles underpin equality law, and there will be many others. The discussion shows how these developing ethical principles enhance the impact of already existing equality law even before there is any AI specific equality legislation. This Report emphasises that Equinet and its Members must engage with the development of such ethical principles both within states and at the level of the EU and CoE. This is a key component of the capacity building work of Equinet and its Members and complements the process of deciding on the next steps to be taken in developing regulation.

Conclusions and recommendations

This Report contains 30 recommendations grouped under three headings. These Recommendations form a comprehensive plan of action for the three main categories of actors involved in regulating the equality implications of AI technologies. These are A) European Equality Bodies themselves, B) states and similar National Authorities, and C) those European bodies that work transnationally across Europe, in particular Equinet, the European Union and the Council of Europe.

A - Recommendations to Equality Bodies

A1	Equinet's Members should designate a team to keep their organisation up-to-date with developments in the AI field. This team should have the primary responsibility to understand the breadth of use of AI systems within each state, their impact on equality, and the ways in which discrimination can occur.
A2	To support this team, or as a part of it, Equinet's Members should consider employing data scientists and other experts to help navigate the complexities of the new technologies.
A3	Equinet's Members should launch public inquiries (or undertake "desktop" reviews of publicly available information) so as to start a process of understanding the ways in which AI is being deployed in their respective territories that potentially impact on the principle of equality and non-discrimination. Regulators and/or academics might be called upon to assist with this exercise. However, as emphasised in Recommendation B1 below, national governments have the primary responsibility for ensuring that there is sufficient transparency in relation to the public uses of AI systems so as to ensure the effective monitoring of AI and the protection of society from the discriminatory impact of AI systems.

A4	Equinet's Members should undertake a legal "gap analysis" to understand how AI systems can be regulated to avoid discrimination and to support equality within their local legal systems, and to identify whether there is a need for local legislative or administrative reform or further Europe-wide legislation.
A5	Equinet's Members should consider the possible need for specific human rights protocols, or new legal forums such as specialist AI courts, to address the equality and non-discrimination issues within their states from AI.
A6	As part of the process outlined in Recommendations A3, A4 and A5, Equinet's Members should review their specific mandates, to ensure that they have adequate and meaningful powers to address the new challenges posed by AI and its challenge to the principle of non-discrimination.
A7	Related to Recommendation A6, Equinet's Members should also identify the financial and logistical resources that they need to undertake the work identified in this Report.
A8	Equinet's Members should use the "gap analysis" referred to in Recommendation No A4 as a "springboard" from which to advance the case for action by their state, for instance, to – • extend the scope of their state's equality, non-discrimination and human rights legislation to cover all protected characteristics and all goods, facilities and services (see Recommendations A4 and A5), • change their mandates to ensure that they are adequate for the tasks that they have specifically identified (see Recommendation A6), and • increase their financial and logistical resources to meet the locally identified challenges (see Recommendation A7).

A9	Equinet's Members should play a leading role in developing and disseminating European and national ethical principles and strategies to guide the implementation of existing laws to address the new challenges posed by AI.
A10	Equinet's Members should provide key information within their states about AI systems and their impact on equality and non-discrimination to individuals, workers, NGOs, businesses, trade unions and even government; they should publish explanatory guides explaining how existing legal provisions can be used to tackle discriminatory algorithms and how AI can be used to the advantage of their communities without causing discrimination.
A11	Equinet's Members should consider undertaking test case and strategic litigation to challenge discriminatory AI systems, both as a means of supporting individuals and so as to make clear that the regulatory enforcement of the principle of equality and non-discrimination will actually happen.
A12	Equinet's Members should initiate and carry out a co-ordinated approach in collaboration with all other relevant regulators because discriminatory AI systems affect many areas, such as finance, data protection, health and safety and product safety, that are within the jurisdiction of other regulators.
A13	Equinet's Members should develop educational and training programmes for organisations, and the public at large, on the human rights and equality impact of AI systems.
A14	Equinet's Members should adopt the checklist set out in Chapter 6 of this Report as a means to ensure that the discriminatory effects of AI systems are identified.
A15	Equinet's Members should engage with academics and similar expert groups to contribute to the development and dissemination of AI related knowledge by the EU and the CoE.

A16 Equinet's members should engage with the faculties of national universities and other academic institutions to ensure that the training of coders includes the understanding of equality.

A17 Likewise Equinet's members should also engage with standardisation initiatives to ensure that European concepts of equality are fully understood and incorporated.

B - Recommendations to states and similar national authorities

B1 National authorities should guarantee greater transparency in the use of AI systems through a comprehensive and systematic mapping of the different ways in which these systems are deployed in their respective territories. The results of such mapping should be made publicly available and should constitute a first step towards ensuring enhanced transparency in the use of AI system. They should develop detailed proposals for the introduction of a legal requirement for transparency through, for example, the creation of a registry for the public uses of AI. This greater transparency should complement and work in parallel with the GDPR which also regulates the use of algorithms but is only a meaningful legal instrument in support of equality where appropriate levels of transparency exist. They should ensure that international trade rules concerning the digital economy do not inhibit transparency.

B2 National authorities in member states of the European Union and the Council of Europe should undertake a legal "gap analysis" to understand how AI systems can be regulated to protect from and prevent breaches of human rights, with due regard to the principle of equality and non-discrimination, and to identify whether there is a need for local legislative or administrative reform or further Europe-wide legislation. They should engage equality bodies in this exercise and should enable them through adequate resources to conduct their own independent legal "gap analysis" focused on the effect of AI systems on equality and non-discrimination. This recommendation sits alongside Recommendation A8.

B3	Further to Recommendation A6, national authorities should support Equinet's Members to review their specific mandates and ensure that they have adequate and meaningful powers to address the new challenges posed by AI.
B4	Further to Recommendation A7, national authorities must ensure that Equinet's Members are adequately and securely resourced to undertake the work identified in this Report. It is up to Equinet's Members to identify the financial and logistical resources that they need for this work.
B5	National authorities should develop and facilitate inter-institutional structures for collaboration and coordination of equality bodies with all other relevant regulators because discriminatory AI systems affect many areas, such as finance, data protection and product safety, that are within the jurisdiction of multiple regulators. This recommendation sits alongside Recommendation A12.
B6	National authorities should ensure that the curriculum for the training of computer scientists, engineers and other professions, concerned with the development of AI systems, includes modules directed to the implications of human rights and equality standards in the development and use of AI systems.

C - Recommendations to Equinet, the European Union and the Council of Europe

C1	Equinet, the EU and the CoE should work together to encourage and facilitate Equinet's Members to be fully aware of the way in which the equality and data protection laws of the European Union and the Council of Europe can operate to control discriminatory and unethical AI systems.
C2	The EU and the CoE should ensure that Equinet and its members are regularly involved in relevant expert groups and legal and policy forums dedicated to the development and dissemination of AI related knowledge in Europe. This recommendation sits alongside Recommendation C1.

C3	Equinet should consider co-ordinating efforts by its Members to undertake specific Europe-wide thematic reviews of the ways in which AI systems are being utilised, for example, one Member could focus on recruitment algorithms, whilst a different organisation might focus on a distinct sector like the financial services industry.
C4	The EU and CoE should work with the national authorities in their respective member states to ensure that the independent oversight over the discriminatory effects of AI systems that Equality Bodies provide is adequately and securely resourced and that equality bodies are equipped with sufficiently broad powers to address the new challenges posed by AI. This recommendation sits alongside Recommendations A6 and A7.
C5	When the EU and CoE consider the possible need for specific human rights protocols or Europe-wide legislation to address the problematic human rights implications of AI systems, they should actively engage and consult equality bodies in these processes. This recommendation sits alongside Recommendation A 8.
C6	The EU and the CoE should encourage and actively support their respective member states to develop educational and training programmes for organisations, and the public at large, on the human rights and equality impact of AI systems, which draw on the expertise of equality bodies.
C7	The EU (and states outside the EU) must ensure that international trade rules concerning the digital economy do not in anyway inhibit the protection of the principle of equality and the elimination of discrimination by making it difficult or impossible to have adequate transparency.

CHAPTER 1: A REVOLUTION THAT AFFECTS US ALL

In the Summer of 2019, when still a candidate for the post of President of the European Commission (EC), Ursula von der Leyen wrote[5] –

> Digital technologies, especially Artificial Intelligence (AI), are transforming the world at an unprecedented speed. They have changed how we communicate, live and work. They have changed our societies and our economies...In my first 100 days in office, I will put forward legislation for a coordinated European approach on the human and ethical implications of Artificial Intelligence. This should also look at how we can use big data for innovations that create wealth for our societies and out businesses.

The Executive Board of Equinet and its General Assembly of Members had correctly anticipated the energy that the new EC would put into this area of work[6] and had seen how the Council of Europe (CoE) also were determined to ensure that these developments were consistent with fundamental human rights.[7] In its 2019 Work Programme, Equinet recognised that across Europe and the globe, algorithms, Artificial Intelligence (AI), machine learning (ML) and automated decision-making (ADM) (collectively in this Report as AI systems) are increasingly, and sometimes stealthily, encroaching upon the space ordinarily inhabited by human actors.

This process has been developing for some years, but now the price and availability of huge computing power has crossed a threshold that makes it possible to undertake computing tasks that, a few years ago, were considered to be impossibly difficult outside a few venues where there were so-called "super-computers."[8] These tasks make AI, ML and ADM, no longer a theoretical possibility but part of the future reality of life for us all.

This change, and the impact it has on human lives, have made it essential that all actors in field of Equality and Human Rights have some understanding of what is involved and what the implications are.

---

5 See "A Union that strives for more My agenda for Europe", see https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf

6 See the work of the High-Level Expert Group on Artificial Intelligence: see https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

7 See the work of the MSI-AUT Committee of experts on the Human Rights Dimensions of automated data processing and different forms of artificial intelligence; see https://www.coe.int/en/web/freedom-expression/msi-aut

8 See the Report of the European Commission's Science and Knowledge Service, Joint Research Centre: "Artificial Intelligence - A European Perspective", 2018 at p.8; see https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf

So, what exactly is the technology with which this Report is concerned? In summary it concerns the use of algorithms. Algorithms are sometimes entirely innocuous, discrete rules that can be followed by a computer; for example, examination boards now frequently use automated systems to mark multiple choice exam sheets. However, algorithms can also be used to make important and nuanced judgements. When algorithms are deployed in this way, there is what is referred to here as "Artificial Intelligence".

There is no single definition of an Artificial Intelligence system. In 2018, the EC adopted[9] a definition of the phrase "Artificial Intelligence" in a communication to the European Parliament, however this was updated by the EU's High-Level Expert Group on Artificial Intelligence (AI HLEG)[10] in 2019 as follows –

> Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.
>
> As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber -physical systems).

Artificial Intelligence systems are thus engaged with data, and often with huge amounts of data. This is what leads to the possibility of apparently intelligent behaviour.

This so-called intelligent behaviour and reasoning is created through ML being the result of the process by which an algorithm analyses data to learn patterns and correlations which are often too subtle, complex and time-consuming for a human to perceive. This process has been helpfully described in some greater detail, by the International Association of Privacy Professionals (IAPP), as follows[11] –

---

9 Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM (2018) 237 final. See https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN

10 See https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341

11 See https://iapp.org/news/a/the-privacy-pros-guide-to-explainability-in-machine-learning/

> Machine learning is a technique that allows algorithms to extract correlations from data with minimal supervision. The goals of machine learning can be quite varied, but they often involve trying to maximize the accuracy of an algorithm's prediction. In machine learning parlance, a particular algorithm is often called a "model," and these models take data as input and output a particular prediction. For example, the input data could be a customer's shopping history and the output could be products that customer is likely to buy in the future. The model makes accurate predictions by attempting to change its internal parameters — the various ways it combines the input data — to maximize its predictive accuracy. These models may have relatively few parameters, or they may have millions that interact in complex, unanticipated ways.

One common application of algorithms and AI is through automated decision-making where conclusions are reached, without any direct, or with only limited, human involvement.

In an age of austerity, AI systems are being used by the public sector to an ever increasing extent.[12] However, it would be wrong to imagine that the growth of AI and ADM is limited to the public sector; the potential commercial benefits of these new types of technology have meant that they are increasingly being embraced by private organisations across Europe.

Beyond the economic benefits, AI systems undoubtedly have enormous potential to further the public good. Recent news stories have described how systems that diagnose cancers such as melanoma and breast cancer, and other diseases such as disorders of the retina, perform better than human experts. The Massachusetts Institute of Technology (MIT) Review[13] recently gave an example of this positive power of AI and ML when describing how an algorithm could perform better than radiologists in diagnosing lung cancer.[14]

Yet there is another side to AI. It can, in certain contexts, if badly designed or incorrectly used, lead to profound breaches of fundamental rights including the principle of non-discrimination. This can happen for reasons such as the use of biased data sets to train the ML algorithm, or by failing to ensure sufficient

12 There is a growing debate concerning the impact on the increasing digitalisation of the state on the poorer parts of communities, see for example, https://chrgj.org/focus-areas/digital-welfare-state-and-human-rights-project/

13 https://www.technologyreview.com/f/613560/google-shows-how-ai-might-detect-lung-cancer-faster-and-more-reliably/ The deep-learning algorithm considered malignant lung nodules in more than 42,000 CT scans. The resulting algorithms turned up 11% fewer false positives and 5% fewer false negatives than human counterparts.

14 Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G. and Naidich, D.P., 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature medicine, 25(6), pp.954-961; see https://www.nature.com/articles/s41591-019-0447-x

transparency and effective human review. Some of these negative outcomes are discussed in greater detail in Chapter 3.

Notwithstanding this scope for harm, a key finding of the research underpinning this Report is that, at present, there is only an embryonic debate across Europe concerning the equality implications of AI systems. While some distinguished experts are considering the interface between equality and these new technologies, the level of knowledge about the work to be done among regulators, NGOs, and other actors in civil society, is still very thin. There is no common minimum level of knowledge amongst Equinet's Members about the nature of AI, ML or ADM nor about the way in which they can lead to discrimination. This is worrying because the lack of this knowledge means that the necessary controls on inappropriate development or use of AI systems are missing or incomplete. Moreover, because discrimination can occur as a result of AI systems in so many different ways, this can be problematic.

There has also been no basis of general common understanding from which specific assessments can be made. That has been an important reason why this Report was commissioned. It is also why Equinet has already been playing its part in raising awareness among Equinet's Members, helping them to understand their new role in this area, considering whether laws to regulate the use of AI, ML and ADM are "fit for purpose", and assessing how discriminatory technology can, and should, be challenged, or utilised to prevent discrimination and promote equality.[15] Moreover Equinet is already undertaking further work such as developing training and this Report will both complement and supplement this work.

There is a further reason for Equinet to be concerned to assist its members to engage with these issues. Much work concerning the development of AI systems is taking place in the United States of America or is backed by capital sourced through the US. There is a well–established practice of group litigation in the US which has concerned many of the US based actors in this field and caused them to consider how best they can protect their systems from chal-

15 The authors do not discuss in this Report the role that AI systems can have in analysing data bases for bias. Nonetheless it is worth recognising that this is one potentially beneficial use to which they have been put; see for instance the work of IBM which has developed its "AI Fairness 360 Open Source Toolkit" – http://aif360.mybluemix.net/?utm_campaign=the_algorithm.unpaid.engagement&utm_source=hs_email&utm_medium=email&utm_content=69523284&_hsenc=p2ANqtz-9vaujms_IQeQkh8nE92xGK7pisSc5eYX3nQkytSKQkCd7rAAd2pPmn_kgreg-FKWVMMD7G0LuVo_jhLB1G1fQZNL81PKA&_hsmi=69523284) and a data set of facial images (https://www.cnbc.com/2019/01/29/ibm-releases-diverse-dataset-to-fight-facial-recognition-bias.html) taken from a Flickr dataset with a 100 million photos and videos with the aim of improving the accuracy, and removing bias from, FRT.

lenges of that sort.[16] As a result, there is a certain amount of work being done in relation to the interface between programming for AI systems and the need to avoid discrimination. However, care must be taken when considering systems developed in the US since they are not always based on a thorough understanding of European equality law, which differs in significant respects from that applied in the US.

US equality law jurisprudence derives largely from the US Supreme Court judgment in 1971 in Griggs v. Duke Power Co.[17] which developed an adverse impact theory. Not long after, in 1978, the U.S. Civil Service Commission, the U.S. Department of Labor, the Equal Employment Opportunity Commission and the U.S. Department of Justice jointly adopted "Uniform Guidelines on Employee Selection Procedures to establish uniform standards for the use of selection procedures by employers and to address adverse impact, validation and record-keeping requirements".[18] These Guidelines use a 4/5ths rule of statistical significance when determining whether there is evidence of a material adverse impact.

On the other hand, the CJEU has not followed this approach. It considered the role of statistical significance in determining whether there is evidence of prima facie indirect discrimination in a number of cases. Most importantly, in Case C-167/97 Regina v Secretary of State for Employment, ex parte Nicole Seymour-Smith and Laura Perez[19] where the UK government argued for a test that was somewhat similar to the Uniform Guidelines, the court held that statistical evidence which reveals "a lesser but persistent and relatively constant disparity over a long period" can establish prima facie indirect discrimination.[20]

When the EU came to make the key equality Directives in 2000, it was made clear that statistical significance was not the only way of establishing prima facie indirect discrimination. For instance, Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation says that prima facie indirect discrimination may"be established by

16 For examples of the kinds of class actions that have been brought in the US see the list compiled by major US law firm Quinn Emmanuel at https://www.quinnemanuel.com/the-firm/publications/article-december-2016-artificial-intelligence-litigation-can-the-law-keep-pace-with-the-rise-of-the-machines/ .

17 401 U.S. 424, 431-2 (1971).

18 See: https://www.govinfo.gov/content/pkg/CFR-2011-title29-vol4/xml/CFR-2011-title29-vol4-part1607.xml

19 ECLI:EU:C:1999:60

20 Ibid, see [60] – [61].

any means including on the basis of statistical evidence",[21] clearly indicating that the EU does not follow the US approach.[22] As well as this difference, the approach to justifying differential treatment is not identical in the US and in Europe, which has developed a particularly rigorous approach.

Of course, the US is carrying out important work in this field, but it proceeds on the basis of the interests of the commercial concepts of equality, and in particular the idea of equality that is dominant in the US. These do not necessarily coincide with the concepts that are part of the corpus of European fundamental rights.

Turning back to this Report, the key questions that have emerged from the research are therefore -

- What tasks should Equinet's Members undertake to ensure that AI, ML and ADM advance and do not hinder equality and non-discrimination?
- What capacity do Equinet's Members have for this?
- How can they be assisted to gain better capacity?
- Who or what are the other actors in this field with which Equinet's Members should be working?
- How well is the European concept of equality supported by the current discourse on the ethical approach to AI and the current regulatory tools?

In the rest of this report, these questions are addressed along with practical recommendations to assist Equinet's Members.

21 See Recital [15]. See also Recital [15] of Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin.

22 See for instance Joined Cases C 159/10 and C 160/10, Fuchs and Köhler v Land Hessen, ECLI:EU:C:2011:508, at [79] and Case C-415/10, Meister v Speech Design Carrier Systems GmbH, ECLI:EU:C:2012:217, at [43] where the CJEU points out how other means might be used.

CHAPTER 2: BROAD USES OF ARTIFICIAL INTELLIGENCE AND RELATED TECHNOLOGIES ACROSS EUROPE

There can no longer be any doubt whatsoever that Equinet and Equinet's Members must have a sufficient degree of understanding of the range of ways in which Artificial Intelligence (AI), Machine Learning (ML) and automated-decision making (ADM) are being used, in order to address the effects of AI. Without that understanding, taking appropriate decisions as to their role, or the interventions that they should be making, whether on behalf of individuals in defending rights or as champions of equality in a wider public discourse, will be challenging or impossible and their prime purpose as organisations will be undermined. The Covid-19 crisis has accelerated the digitalisation of services and the collection of sensitive data and increased their potential to impact on all kinds of equality issues, for instance with a rapid increase with use of contact tracking devices.[23]

To this end, this chapter maps some of the most common and significant ways in which this technology is being deployed. This is based on the information, collected from research, particularly from AlgorithmWatch,[24] or that has been supplied by Equinet's Members. From this, it is clear that AI systems, including ML, algorithms and ADM, are being used across Europe in a multitude of different ways.

Appendix 1 sets out the collated information concerning the spread of these technologies within Equinet's Membership in so far as this information is publicly available.[25] It will be seen that the examples in Appendix 1 cover a very large range of contexts, and there is a sound basis to infer that it is very likely that where it is known that AI, ML and ADM is being used in a particular context in one country, it is very likely that it is already being used or soon will be used, in a similar context in other countries. Equinet's Members would therefore be well advised to consider the information in Appendix 1 as indicating not merely what is happening in their country, but what is likely to be happening in the near future as well.

23 See for instance Clift K., Court A., "How are companies responding to the coronavirus crisis?", World Economic Forum, 23 Mar 2020; see https://www.weforum.org/agenda/2020/03/how-are-companies-responding-to-the-coronavirus-crisis-d15bed6137

24 https://algorithmwatch.org/en/

25 This information has been collected during 2019 and there is no doubt that by the time of the publication of this Report it will be out of date, as the uses in one state will be adopted or applied in others, and as new uses are developed.

Appendix 1 identifies some common uses of AI as follows -

- Reducing unemployment: Algorithms are being used at least in Austria, Spain, Sweden and Poland to calibrate the level of assistance which jobseekers should receive from the state in relation to obtaining employment. Equally, in Belgium, algorithms are being deployed to assess the extent to which jobseekers are actively engaging in attempts to secure employment with the potential threat of sanctions for those who are insufficiently dedicated.
- Facial recognition technology (FRT) and other forms of biometric identification technology: Biometric identification technology dependent on ML algorithms, is being deployed by state actors at least in France, Italy, Sweden and the UK. AlgorithmWatch currently estimates that at least 10 police forces across Europe are using FRT.[26]
- Education: Algorithms are being used in a variety of ways in France, Italy, Poland, Slovenia and Sweden from the allocation of teachers to the monitoring of students.
- Recruitment: Complex AI is being used in Finland to assess candidates for roles including through automated video analysis and assessment of social media presence. Companies are also deploying technology in the recruitment field in at least France, the Netherlands, Sweden and the UK. There is a fast-developing commercial interest in the use of AI systems to assist with recruitment in the US,[27] and some of these will undoubtedly be used here in Europe.
- Predictive policing: Sophisticated AI is being used to make predictions about where crimes will be committed and by whom in Belgium, Germany, the Netherlands, Spain and the UK.
- Immigration / border control: AI is being used in at least Slovenia and the UK to make decisions about immigration status and to control borders.
- Financial products: Algorithms are being used in relation to credit scoring and the availability of insurance in Denmark, Finland, France, Germany, Slovenia, Sweden and the UK.
- Health: Technology is being used to detect and predict illnesses in Denmark, France, Italy, Spain, Sweden and the UK.

---

26 See https://algorithmwatch.org/en/story/face-recognition-police-europe/

27 This has already led to the State of Illinois introducing an Artificial Intelligence Video Interview Act 2019 (see http://www.ilga.gov/legislation/publicacts/fulltext.asp?Name=101-0260) to regulate the practices of firms such as HireVue that use AI systems to determine through video interviews who should be recruited.

- Social advantages: Algorithms are being utilised to make decisions concerning eligibility for social welfare in countries such as Estonia, Finland, the Netherlands, Sweden and the UK; these governments have adopted these processes in order both to reduce the cost of determination and to try to increase predictability, so it is to be expected that other countries will also wish to adopt similar processes.
- Child welfare: Some countries like Spain and the UK are deploying AI to assess the risk of children requiring state interventions in order to protect their welfare.
- Justice and criminal justice system: Countries where AI is being used or contemplated in the justice and criminal system include France, Italy, the Netherlands, Poland, Spain and the UK.
- Fraud detection: AI is being used to predict which individuals might be defrauding the state in Italy, the Netherlands, Poland, France, Sweden and Slovenia.
- Military systems: Many countries also consider that AI, ML and ADM are significant for the control of public order through military systems. In December 2019, the European Council on Foreign Relations published a Policy Brief emphasising that European countries and the EU will soon have to engage with the potential for AI, ML and ADM to enhance its military capabilities and for the regulation appropriate to this next step.[28] While this kind of use will often be concerned with external threats it is likely that it will also be used, on occasion, internally.

It seems very likely that the AI systems identified above are being used across Europe both by state actors, and also by the commercial world, even where that information is not publicly available.

The picture in Europe appears to be very much consistent with broader global trends. Whilst it is impossible to calibrate precisely the extent to which AI is expanding, there are some resources which provide fleeting insights. Stanford University has produced an online "AI Index" which tracks the growth of AI by numerous metrics including country, sector and even business function.[29] This Index demonstrates an explosion in AI in the past 2 to 3 years. Indeed, in

28 See Ulrike Esther Franke, "Not smart enough: the poverty of European military thinking on Artificial intelligence", European Council on Foreign Relations, see https://www.ecfr.eu/page/-/Ulrike_Franke_not_smart_enough_AI.pdf

29 Raymond Perrault, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz, Terah Lyons, James Manyika, Saurabh Mishra, and Juan Carlos Niebles, "The AI Index 2019 Annual Report", AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, December 2019. Available at https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf

September 2019, KPMG released its report, "KPMG 2019 Enterprise AI Adoption Study into AI" which found that 30% of thirty of the world's largest companies, with aggregate revenues of $3 trillion, were deploying AI while 17% said they have deployed AI and ML at scale across their enterprise.[30]

The drive to accelerate the use of these technologies is immense. Indeed, in February 2019 the McKinsey Global Institute calculated that there is a huge potential gain for Europe saying that[31] -

> If Europe on average develops and diffuses AI according to its current assets and digital position relative to the world, it could add some €2.7 trillion, or 20 percent, to its combined economic output by 2030. If Europe were to catch up with the US AI frontier, a total of €3.6 trillion could be added to collective GDP in this period.

However, whether this prediction is wholly accurate does not really matter. It reflects the determination of the EU to move rapidly forward in developing AI technologies as set out in the "Coordinated plan on artificial intelligence" published at the same time.[32]

The EC has reiterated the expected benefits of such development in the White Paper "On Artificial Intelligence - A European approach to excellence and trust" published in February 2020,33 emphasising that it "...supports a regulatory and investment oriented approach with the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of this new technology."

The breadth of the use of AI as identified in this chapter and in Appendix 1 emphasises just why Equinet's Members cannot afford to be ignorant of these developments. In short it is already critical that their managing boards or committees understand generally, and that at least a team within each member understands more particularly, the breadth of this use of AI systems. Without this background knowledge it will be impossible for Equinet's Members to begin to consider the ways in which these new forms of technology interplay with the fundamental principle of non-discrimination.

Therefore, there is a widespread need for training on these issues, and it will be necessary for Equinet's Members to have a designated person or, preferably, a

30 See https://advisory.kpmg.us/articles/2019/ai-transforming-enterprise.html

31 See https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-europes-gap-in-digital-and-ai

32 See https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/en/pdf

33 See https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

team, responsible for keeping the organisation up to date with developments as they occur.

CHAPTER 3: ARTIFICIAL INTELLIGENCE WITHIN EUROPE'S EXISTING EQUALITY AND DATA PROTECTION LAWS

Equality laws

If Equinet's Members are to begin undertaking the task of regulating the equality impact of these new AI systems, they will need to understand what legal resources they have at their disposal.

While at present, there is no legislation which is specifically directed to prohibiting discriminatory AI systems within European law,[34] there is an already well-established framework of laws and international standards for the protection and advancement of equality and non-discrimination which can be drawn on. There is no doubt that in certain circumstances, AI, ML, algorithms, and ADM will infringe Europe's existing equality framework.

There is range of documents and reports that can assist with understanding what can be done with existing laws. For instance, the CoE, the EC, and the Fundamental Rights Agency of the European Union (FRA), have all produced proposals or analyses on the interplay between AI and fundamental rights, such as the principle of non-discrimination. An account of these initiatives is set out in Appendix 2 to which Equinet's Members should refer for a more detailed analysis as they develop their capacities to address potentially discriminatory AI systems. In this chapter, some of the typical ways in which AI systems can be discriminatory and offend existing equality laws are demonstrated.

Within the EU, Equinet's Members will be able to draw on the laws and principles that form part of the acquis communautaire.[35] Outside the EU, Equinet's Members will be able to look to the jurisprudence developed within the CoE, and the more general international law instruments deriving from Universal Declaration of Human Rights and other UN Conventions. While the aim of

34 There have already been calls for changes to the law to provide for this; see for instance https://datenethikkommission.de/wp-content/uploads/191023_DEK_Kurzfassung_en_bf.pdf This issue has also been raised by the European Commission in its white paper on future proposals to regulate AI "On Artificial Intelligence – A European approach to excellence and trust". See https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf These are discussed further below.

35 That is to say a body of common rights and obligations that are binding on all EU countries as EU Members: https://eur-lex.europa.eu/summary/glossary/acquis.html

this chapter is to explain how these laws can impact on AI, ML and ADM, it is emphasised that this is not intended to provide a wholly comprehensive survey, but through examples of the kinds of discrimination that can occur when AI, ML and ADM is not used appropriately, to demonstrate the possibilities that already exist for intervention by Equinet's Members.

Many states within the EU and/or within the CoE will also have well-developed domestic equality laws, which will often have a further reach.[36] Obviously, Equinet's Members can seize the opportunities that such domestic laws might offer for further and deeper interventions, but a detailed examination of these local laws is outside the scope of this report. What is presently available to Equinet's Members to undertake this new role regulating the equality impact of AI systems will now be considered.

The Amsterdam Treaty brought Article 13 (now Article 19 TFEU) into the EC Treaty permitting the Council to make non-discrimination laws on a wide range of grounds: sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation.[37] Thereafter, the EU approved two major equality Directives[38] and recast the legislation prohibiting discrimination on grounds of gender,[39] as well as adopting the Charter of Fundamental Rights of the European Union (CFREU),[40] as a substantive provision of the Lisbon Treaty.[41] The Court of Justice (CJEU)

36 In the UK, for example, the Equality Act 2010 is a powerful tool to challenge discrimination in the employment field and also the provision of goods, facilities and services across the private and public sector plus clubs and associations.

37 Article 19 TFEU now says "1. Without prejudice to the other provisions of the Treaties and within the limits of the powers conferred by them upon the Union, the Council, acting unanimously in accordance with a special legislative procedure and after obtaining the consent of the European Parliament, may take appropriate action to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation. 2. By way of derogation from paragraph 1, the European Parliament and the Council, acting in accordance with the ordinary legislative procedure, may adopt the basic principles of Union incentive measures, excluding any harmonisation of the laws and regulations of the Member States, to support action taken by the Member States in order to contribute to the achievement of the objectives referred to in paragraph 1. "; see https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:12008E019:EN:HTML

38 Directive 2000/43/EC (6) implementing the principle of equal treatment between persons irrespective of racial or ethnic origin and Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation.

39 Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast).

40 See http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

41 See https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:ai0033&from=EN

has also increasingly positively stated that equal treatment is a general or fundamental principle on which the EU is founded.[42]

This EU body of law also draws on the equality law jurisprudence of the European Court of Human Rights (ECtHR), which was originally[43] derived from Article 14 of the CoE's European Convention on Human Rights (ECHR),[44] and relevant United Nations Conventions, such as the UN Convention on the Rights of Persons with Disabilities.[45] All Equinet's Members ought to be able to use this jurisprudence.

In summary, European laws provide a very wide-ranging protection in relation to employment and occupation and in the provision of goods, facilities and services and the provision of social advantages in relation to discrimination on grounds of race and ethnicity and gender. The protection is less extensive in relation to religion or belief, disability, age or sexual orientation, which is centred on employment and occupation only. The protections in the ECHR are more general in the range of characteristics protected, and more general in their application.

Nonetheless, it would be a mistake to think that this difference in the level of protection is mirrored across Europe at a national level within domestic legislation. Many states have a coherent set of equality law rights for each of the protected characteristics in both the employment field and the provision of goods, facilities and services. The general approach of international conventions is to require this, and in any event, it is usually demanded by civil society. There are several good websites which discuss the comparative protections from discrimination across Europe, and which demonstrate the local implementation across the full range of circumstances.[46]

The principle of non-discrimination as applied to Artificial Intelligence

This Report considers, in summary, how AI, and associated forms of technology, could infringe the basic tenets of discrimination law in Europe (both in EU member states and those states not in the EU which are members of the CoE).

---

42 See for instance Case C-144/04, Werner Mangold v. Rüdiger ECLI:EU:C:2005:709.

43 Article 14 ECHR is well known to be a so-called parasitic provision which can only have effect within the sphere of application of another provision of the ECHR. Protocol 12 to the ECHR (see https://www.echr.coe.int/Documents/Library_Collection_P12_ETS177E_ENG.pdf) replicates the terms of Article 14 but is not so limited and has general application. Not all Member States have given full effect to this Protocol, but it should also be noted that some states such as Estonia, the Netherlands, Poland and Finland have open-ended non-discrimination provisions in their constitutions.

44 See https://www.echr.coe.int/Documents/Convention_ENG.pdf

45 See for instance Case C-395/15 Daouidi ECLI:EU:C:2016:917.

46 See for instance https://www.equalitylaw.eu/publications/comparative-analyses

This is achieved by examining four case studies[47] relating to (i) a basic ADM algorithm, (ii) FRT, (iii) sophisticated algorithms that predict risk in the context of social welfare, and (iv) algorithms that assist the government, in our example the UK, to determine immigration status.[48]

Case Study A: Basic automated decision-making algorithm

Some organisations will use algorithms to speed up decision making. These are not necessarily sophisticated, but they can become infected with discrimination, either through ML based on inappropriate data sets, or simply because they reflect the prejudice of the coder who designed the algorithm which has been used on a data set. The following simple example is given, relating to a chain of gyms in the UK, to demonstrate how this can happen.

A woman was a member of a well-known commercial gym company, with many venues across the UK. She was also a Doctor of Medicine by profession. A problem occurred when she was unable to use a swipe card provided by the gym company in order to access locker rooms at one of its gym venues. The problem was investigated and it transpired that the gym company was using a computer system that used a member's title to determine which changing room (male or female) a gym customer would be permitted to access.[49] The computer system had an algorithm that searched the database of the gym company's members, to identify their gender and then to allocate permissions in accordance with that assessment. The aim was simple: to ensure that women went to the female changing rooms and men to the male changing rooms. The algorithm used by the computer determined gender and therefore access by reference to the gym member's title. The problem was that the algorithm identified "Doctor" as a "male" identifier. Accordingly, this female doctor was not permitted by the computer system to enter the women's changing rooms.

This case provides a very simple, and indeed classic, example of direct sex discrimination. The customer was treated less favourably because she was a woman in circumstances in which a comparable male doctor would not have been. Because European law does not permit direct sex discrimination of this kind ever to be justified, the gym had to recognise immediately that it had made

47 For additional case studies, see https://ai-lawhub.com

48 The authors are grateful to Swee Leng Harris of the Legal Education Foundation in the UK for bringing to their attention case studies two and three, both of which form the basis for their legal opinion at https://www.cloisters.com/wp-content/uploads/2019/10/Open-opinion-pdf-version-1.pdf

49 See https://metro.co.uk/2015/03/18/gyms-computer-assumed-this-woman-was-a-man-because-she-is-a-doctor-5110391/

a mistake and would be liable to her. Fortunately, it had the good sense to acknowledge its fault and to make reparation without the need for any litigation or further intervention. However, that might well not have been the case and it is easy to imagine similar scenarios in which this did not happen.

This is a simple case to understand, but a more difficult issue is when facial recognition technology (FRT) discriminates, which is considered next.

Case Study B: Facial recognition technology (FRT)

There are many forms of biometric identification which use AI and ML. However, the one most discussed is FRT. As humans, most of us learn to navigate the world through facial recognition from our very earliest days on earth. Faces are personal identifiers and as such are of great interest to the new technologies. Issues arise when facial images are changed, as with proprietary software available on Instagram or through Deepfake programmes which involve the digital "undressing" of a person or the substitution of one person's face for that of another.

These can easily lead to harassment, including harassment on grounds which are protected by equality laws. Equinet's Members need to be aware of these issues, however, here the Report focuses more closely on the kinds of FRT which is increasingly being used to verify identities as a gateway to access or deny access to a range of goods facilities and services. Individuals may be most aware of these, when for instance they use the electronic gates at a border frontier and when they are required to stand still and be knowingly scanned, but FRT is used in a myriad of other situations, when frequently people are not fully aware how it is monitoring us.

FRT is becoming increasingly cheap to purchase and so its deployment is increasing all the time. Many well–known companies such as Amazon,[50] IBM,[51] and Microsoft[52] have proprietary products; in some circumstances access to these is

50 For instance see https://aws.amazon.com/free/machine-learning/?trk=ps_a131L-0000057ji8QAA&trkCampaign=acq_paid_search&sc_channel=ps&sc_campaign=acquisition_uk&sc_publisher=google&sc_category=Machine%20Learning&sc_country=UK&sc_geo=EMEA&sc_outcome=acq&sc_detail=%2Bfacial%20%2Brecognition&sc_content=facial_recognition_bmm&sc_segment=377966061761&sc_medium=ACQ-P|PS-GO|Non-Brand|Desktop|SU|Machine%20Learning|Solution|UK|EN|Text&s_kwcid=AL!4422!3!377966061761!b!!g!!%2Bfacial%20%2Brecognition&ef_id=EAIaIQobChMIr_uG_dXx5gIVyLHtCh3Jew34EAAYASAAEgLbFfD_BwE:G:s

51 For instance see https://cloud.ibm.com/catalog/services/visual-recognition

52 For instance see https://azure.microsoft.com/en-gb/services/cognitive-services/face/

free. Other companies offer dedicated FRT products promising a very high level of utility in specific circumstances.

In the context of a border-crossing, the FRT system usually requires the traveller to remove spectacles, to look straight at the camera and not to smile. The computer then makes a match against a known photograph which has similarly been taken without spectacles or a smile and which is a full-face image. FRT systems are being developed that do not require such a careful presentation of the face and which seek to match faces to known images even when the FRT is confronted with very different angles of view.

While the effectiveness of FRT in such contexts is still keenly debated, it is already being widely deployed across Europe (see Appendix 1). Equinet's Members should be aware that these systems will provide false matches or sometimes fail to make matches when they would be appropriate. These are false positives and false negatives. It is well established that they can occur on a discriminatory basis and that this depends on the competence of the AI system to make appropriate matches. This skill in the system is learnt by the computer as a result of ML using databases of already identified faces.

Research in the US by Joy Buolamwini and Timnit Gebru revealed how in the US this type of technology can have a disparate impact on women and certain racial groups.[53] They highlighted how commercially available systems contained a misclassification error rate of up to 34.7% for darker skinned women in comparison to a maximum error rate of 0.8% for lighter skinned males. It is obvious that if such a faulty FRT system were to be used in Europe as a gateway to a benefit or service of some kind it would be potentially discriminatory. The Report of Buolamwini and Gebru, and other researchers' work has prompted much analysis of the problem and consideration as to how FRT systems can be improved. Nobody doubts that this will happen but there is still a very long way to go.

Thus, a Report published by the United States Department of Commerce's National Institute of Standards and Technology in December 2019 concluded that there were still many problems with widely accessible FRT products particularly in relation to false positives.[54] It noted that –

---

53 Buolamwini, J. and Gebru, T., 2018, January. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency, PMLR 81:77-91, 2018; see http://proceedings.mlr.press/v81/buolamwini18a.html

54 Grother, P., Ngan, M. and Hanaoka, K., 2019. Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. National Institute of Standards and Technology; see https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

> ... false positive differentials are much larger than those related to false negatives and exist broadly, across many, but not all, algorithms tested. Across demographics,
>
> false positives rates often vary by factors of 10 to beyond 100 times. False negatives tend to be more algorithm-specific and vary often by factors below 3.

False positives are likely to be particularly important from the point of view of equality and non-discrimination since they are more likely to lead to adverse and unjustified interventions. It is therefore important that the Report also concluded that these false positives were related to the place of use, finding that[55] –

> ...false positive rates are highest in West and East African and East Asian people, and lowest in Eastern European individuals. This effect is generally large, with a factor of 100 more false positives between countries. However, with a number of algorithms developed in China this effect is reversed, with low false positive rates on East Asian faces. With domestic law enforcement images, the highest false positives are in American Indians, with elevated rates in African American and Asian populations; the relative ordering depends on sex and varies with algorithm. We found false positives to be higher in women than men, and this is consistent across algorithms and datasets. This effect is smaller than that due to race. We found elevated false positives in the elderly and in children; the effects were larger in the oldest and youngest, and smallest in middle-aged adults.

The authors argue that there is no evidence to suggest that the forms of FRT deployed within Europe would be any better than this example. The use of FRT in Europe could readily give rise to indirect discrimination. Indirect discrimination occurs where an apparently neutral provision (here, the algorithm or the data used to train the algorithm) puts or would put persons with a protected characteristic (for instance, ethnicity and/or gender) at a particular disadvantage (here, the risk of being misidentified) compared with others (different gender/ different ethnicity).[56]

In such a scenario, there will be unlawful discrimination unless the FRT can be objectively justified by reference to a legitimate aim, and even then, it will only be justified if the means of achieving that aim are appropriate and necessary. Where the user of the FRT is a public body[57] the jurisprudence of the ECtHR will entail asking whether this use is appropriate and necessary within a democratic context in accordance with the case law of Article 14 ECHR.

---

55 Ibid.

56 See for instance the Article 2(1)(b) of Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast).

57 This is likely also to be the case if the user is a private company too.

More broadly, it can be said that there are three key hurdles that must be crossed before a justification defence would be successful; these are that -

- the measure adopted by the service provider is underpinned by a legitimate aim;
- the measure is capable of achieving that aim;
- and the measure is proportionate.[58]

Importantly, a measure will not be proportionate where the aim could be achieved through a different measure which was less discriminatory or not discriminatory at all.

In many contexts an organisation deploying FRT could have a legitimate aim for its use. These might include seeking to identify individuals quickly and accurately. Yet it may face real problems when it comes to showing that such an aim was being achieved by the FRT in question. It may have an even bigger problem in showing that the aim was being achieved in a proportionate way.

This is for two reasons –

- As noted, much research shows that FRT does not accurately classify people. This is not just a problem in the US or in China. Independent research published by the University of Essex into the activities of the Metropolitan Police Service in London noted that FRT had a poor record of assisting the police in accurately identifying individuals.[59] Specifically, across test deployments, 63.64% were verified incorrect matches and only 36.36% were verified correct matches. If the FRT in question had such a low success rate, it can hardly be said that it is achieving its aim of seeking to accurately identify people. It is considered that any justifi

  cation defence in relation to the use of this system would fail because it can hardly be said that its aim is being achieved.
- Secondly, it is known that FRT can be made "less biased" by simply training it on better data. Indeed, as part of their research Buolamwini and Gebru, sought to cure the bias they had identified by creating a new data set based on a more balanced representation of both gender and racial diversity, drawn from the members of the national assemblies of a very wide number of different countries and using a better mix of genders. Using this data set, the researchers found that by training the FRT on a

58 Case C-170/84 Bilka-Kaufhaus GmbH v Weber von Hartz ECLI:EU:C:1986:204.

59 Fussey, P. and Murray, D., 2019. Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology; see https://48ba3m4eh2bf2sksp43rq8kk-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report.pdf

non- (or at least much less) biased selection of faces the AI system was much more successful. The message of this research is that users of FRT must train their systems on non–discriminatory data sets otherwise they will not be able to show that the use of the FRT was a proportionate means of achieving any legitimate aim. Put another way, if FRT is potentially indirectly discriminatory, it is hard to see how it could ever be justified if there was a better system potentially available, as the US Department of Commerce research shows will often be the case.

Moreover, in the AI sphere, where there are real concerns as to the ethics of ADM, it is suspected that a body of jurisprudence will develop that will be heavily influenced by the emerging debate over "ethical AI". A comprehensive inventory of the various ethical principles which have been published across the globe is maintained by Algorithm Watch.[60]

The leading statement within the EU on the ethical use of AI are the draft "Ethics Guidelines for Trustworthy Artificial Intelligence (AI)". These Guidelines were prepared after extensive consultation by the AI HLEG,[61] and are discussed further in Appendix 2. They heavily emphasise that AI must only be used to improve collective and individual well-being. The over-arching theme is that AI systems must be "human-centric", through working to the standards set by the AI HLEG, thus –

> The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. This also entails consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come

So, it is very likely that an aim will only be legitimate for the purpose of an objective justification defence under European law insofar as the AI system is intended to achieve this aim of improving collective and individual well-being. Accordingly, it is very likely that FRT that gives rise to prima facie indirect discrimination, will only be justifiable insofar as it also promotes collective and individual well-being. There will be contexts in which FRT will be deployed in such a way, for example, where it leads to improvements in personal safety. However,

60 It is available here: https://inventory.algorithmwatch.org/ and demonstrates the sheer array of ideas in this area.

61 https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

the use of facial recognition in more mundane commercial contexts may well be incapable of justification if the law develops in the direction anticipated.

The paper, "The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)" will be highly relevant to the question of proportionality, which is again a facet of the objective justification test. These have been highlighted in Appendix 2. In addition to identifying the purpose of "ethical AI", this paper explains the principles which should define any AI system as follows: respect for human autonomy, prevention of harm, fairness, and explicability. It is expected that these principles will also shape the future discussion concerning proportionality within the justification defence. It is suspected that unless an organisation can show that their AI systems comply with these ethical principles, then they will be unable to satisfy the test of objective justification.

The ethical principle of "explainability" is particularly important as many organisations will struggle to demonstrate that their AI systems are transparent due to the "black box" problem; the AI HLEG have noted that[62] –

> Explainability concerns the ability to explain both the technical processes of an AI system and the related human decisions (e.g. application areas of a system). Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. Moreover, trade-offs might have to be made between enhancing a system's explainability (which may reduce its accuracy) or increasing its accuracy (at the cost of explainability). Whenever an AI system has a significant impact on people's lives, it should be possible to demand a suitable explanation of the AI system's decision-making process. Such explanation should be timely and adapted to the expertise of the stakeholder concerned (e.g. layperson, regulator or researcher). In addition, explanations of the degree to which an AI system influences and shapes the organisational decision-making process, design choices of the system, and the rationale for deploying it, should be available (hence ensuring business model transparency).

This problem arises because in many cases it is impossible to look inside an algorithm, AI or ML process, to understand how decisions are being made. The AI HLEG have described this problem thus[63] –

> Black-box AI and explainability.
>
> Some machine learning techniques, although very successful from the accuracy point of view, are very opaque in terms of understanding how they make decisions. The notion of black-box AI refers to such scenarios, where it is not possible to trace back to the reason for certain decisions. Explainability is a property of those AI systems that instead can provide a form of explanation for their actions.

62 See https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1

63 See https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651

Jenny Burrell, an academic who specialises in AI, has also very simply made the point that[64] –

> While datasets may be extremely large but possible to comprehend and code may be written with clarity, the interplay between the two in the mechanism of the algorithm is what yields the complexity (and thus opacity).

In so far as an AI system is not sufficiently transparent and it leads to prima facie discrimination, it is predicted that organisations will encounter great difficulties in proving objective justification.

Indeed, it is very possible that the lack of transparency itself will lead to the courts finding that the technology is prima facie discriminatory. That is, in equality law it is well established that a lack of transparency in a pay system can give rise to an inference of discrimination. This was established some thirty years ago in Case C-109/88 Danfoss[65] and has been reiterated on many occasions. There is no reason why this principle would not extend to AI. So, paradoxically, the lack of meaningful transparency as to the way in which an algorithm or AI or ML works, might assist claimants or organisations who are challenging technology which might be discriminatory.

These are issues of process for equality law, but it should be recognised that they may also have substantive consequences in terms of the General Data Protection Regulation (GDPR).[66] The relevance of the GDPR is considered below in Chapter 3.

64 Burrell, J., 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), p.2053951715622512; see https://journals.sagepub.com/doi/10.1177/2053951715622512

65 Case C- 109/88, Handels- og Kontorfunktionaerernes Forbund i Danmark v Dansk Arbejdsgiverforening Ex p. Danfoss A/S ECLI:EU:C:1989:383.

66 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) (Text with EEA relevance).

Case Study C: Predicting risk

Many of the countries from which Equinet's Membership is drawn are using AI systems to predict the risk of a certain occurrence (see Appendix 1). These include the following assessments –

- the risk of a person remaining unemployed,
- the risk of an elderly person requiring care,
- the risk that a child might need welfare services,
- the risk of a crime,
- the risk of hospitalisation,
- the risk of committing fraud and
- the risk of re-offending.

Risk analysis is a key area where discrimination can occur in a way which can have significant effects on individuals. To exemplify this point, this Report analyses one system of predicting risk used in the UK called "Risk-Based Verification" (RBV) within an equalities framework.

In the UK, local authorities are required under legislation to determine an individual's eligibility for Housing Benefits and Council Tax Benefits. There is no fixed verification process but local authorities can ask for documentation and information from any applicant *"as may reasonably be required"*.[67] Since 2012, the Department for Work and Pensions (DWP) has allowed local authorities to voluntarily adopt RBV systems as part of this verification process for applications and has given guidance as to how this may happen.[68]

It is understood that RBV works by assigning a risk rating to each applicant for Housing Benefit and Council Tax Benefit, which then determines the level of identity verification required. This allows the local authority to target and focus resources on *"... those cases deemed to be at highest risk of involving fraud and/or error"*.[69] For example, an individual with a low risk might simply need to provide proof of identity but someone with a high-risk rating might be subject to Credit Reference Agency checks, visits, increased documentation requirements etc.[70]

---

67 The Council Tax Benefit Regulations 2006, SI 2006 No. 215, reg 72; see http://www.legislation.gov.uk/uksi/2006/215/regulation/72/made and the Housing Benefit Regulations 2006, SI 2006 No. 213 reg 86:: http://www.legislation.gov.uk/uksi/2006/213/regulation/86/made

68 Housing Benefit and Council Tax Benefit Circular, HB/CTB S11/2011; see https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633018/s11-2011.pdf

69 Ibid.

70 Ibid.

A DWP circular shows that the Department is aware that ML algorithms are being deployed as part of this process.[71] However, it has been impossible to identify any publicly available information that explains *how* such algorithms are being deployed, or *on what basis*.

That aside, there is good reason to believe that the use of RBV may well give rise to discrimination in some instances. For example, an audit noted the high degree of false positives, that the ML algorithm consistently detected a far greater percentage of "*high risk*" applicants than had been anticipated[72] -

Year	Detection vs Expectation
2015/2016	33% vs 20%
2016/2017	33% vs 20%
2017/2018	33% vs 20%

When a random sample of 10 of the "*high risk*" applicants was further examined, those on the list were all found to be women who were working. This could be a coincidence, as the sample was small, or it could suggest that the algorithm had "*learnt*" a discriminatory correlation. It ought to have rung alarm bells since it is well-established from studies of AI that pattern recognition technology can unintentionally lead to the replication of human biases in various subtle ways. For instance, the UK's House of Commons Science and Technology Select Committee noted this in 2018, pointing out how ML algorithms can, far from introducing objectivity, actually perpetrate discrimination through learning discriminatory relationships between data.[73]

Accordingly, it is possible that the RBV systems utilised in the UK or the myriad of other AI systems in use across Europe which predict "*risk*" could be acting in a discriminatory way. However, because of the "*black box*" problem described above, it is very difficult to understand precisely what is happening so as to ensure that technology is being deployed in a way which is free from discrimination. Accordingly, it is anticipated that AI systems which predict risk, but which

71 Ibid.

72 Ibid.

73 "Algorithms in decision-making", House of Commons Science and Technology Committee Fourth Report of Session 2017–19 Report, 15 May 2018, HC 351; see https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf

cannot be examined transparently, are very likely to be litigated in the future with litigants relying on the principle in *Danfoss* that a lack of transparency can give rise to an *inference* of discrimination.

Case Study D: Immigration status

Algorithms are being deployed in Europe in relation to immigration decision-making and determining whom may claim citizenship (see Appendix 1).

One well publicised use of such technology is the Settled Status scheme, established by the Home Office because of the extra demands made by Brexit, and used in the UK to regularise the immigration status of EU, European Economic Area (EEA), Swiss nationals and their families living in the UK.[74] Success in this process is very valuable, permitting an individual to remain in the UK after 30 June 2021.

Settled Status is ordinarily awarded to qualifying individuals who started living in the UK by 31 December 2020 (or by the date the UK leaves the EU without a deal) and who have lived in the UK for a continuous five-year period (known as '*continuous residence*').[75] Five years' continuous residence means that for five years in a row an individual has been in the UK,[76] for at least six months in any twelve-month period.

In order to determine if an individual has been resident for the relevant five year continuous period, the Home Office application process will often use automated data processing to analyse data from the DWP and the tax authorities (referred to as the HMRC) to verify how long an individual has been in the UK.[77]

The benefits data held by the DWP, which is examined by the algorithm, consists of thirteen categories: State Pension and New State Pension, Housing Benefit, Jobseekers Allowance, Employment Support Allowance, Carers Allowance, Universal Credit, Personal Independent Payment, Disability Living Allowance, Income Support, Maternity Allowance, Incapacity Benefit, Attendance Allowance and Severe Disablement Allowance.[78]

The precise way in which the ADM algorithm reaches the conclusion that an

74 See https://www.gov.uk/settled-status-eu-citizens-families

75 See https://www.gov.uk/settled-status-eu-citizens-families/what-settled-and-presettled-status-means

76 Or the Channel Islands or the Isle of Man.

77 See https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check

78 See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790668/Home_Office_-_DWP_API_EU_Exit_MoU.PDF

individual has been resident during certain periods of time is not entirely clear. However, it does appear that a case-worker will use the data provided by the algorithm, which is simply the months that the algorithm has determined the applicant was resident, to "*inform a calculation to determine whether an applicant's UK residence indicates whether they are eligible for consideration*" under the scheme.[79] It also appears that the case-worker will be able to exercise some discretion when reaching a decision but when, how and on what basis that discretion is exercised is unclear.[80]

There is also a stage within the process whereby individuals who have entered the ADM process, can upload documentation in order to verify periods of residence that could not be confirmed by the algorithm.[81] Importantly, however, an applicant will *not* be informed of the reason that an automated check has concluded that the person does not have continuous residence during a certain period. The rationale for this system is that "*... because doing so may introduce the risk of identity theft and abuse*".[82] Accordingly, another layer of opacity is introduced into the system.

Finally, there is a right to seek an administrative review of the decision reached by the case-worker although, inevitably, any review would be lodged in ignorance of the precise problem with the application.[83]

The government's aim when implementing a system which relies so heavily on an automated algorithm are as follows[84]: reducing reliance on paper documentation, reducing caseworker processing time, reducing fraud and error, improving customer experience, and minimizing the evidential burden on applicants, especially in light of an anticipated "*influx*" of applications. Evidence[85] suggests

---

79 See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790668/Home_Office_-_DWP_API_EU_Exit_MoU.PDF

80 See https://www.childrenslegalcentre.com/wp-content/uploads/2019/03/EUSS-briefing_Mar2019_FINAL.pdf

81 See https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check

82 See https://www.gov.uk/guidance/eu-settlement-scheme-uk-tax-and-benefits-records-automated-check

83 See https://www.gov.uk/guidance/eu-settlement-scheme-apply-for-an-administrative-review

84 See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790668/Home_Office_-_DWP_API_EU_Exit_MoU.PDF. The equivalent document for the HMRC contains the same information and is available here: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790661/Home_Office_-_HMRC_API_EU_Exit_MoU.PDF https://www.gov.uk/government/publications/eu-settlement-scheme-private-beta-2

85 See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_

that the government has been successful in achieving some of these aims, for example –

- 75% of applications using the app were able to prove their identity in under 10 minutes.
- 79% of applicants found the process "*very or fairly easy*".
- 73% of applicants did not need to submit additional data to prove residence after the algorithmic review process.

However, there is a possible different perspective here, since, as further explained below, there is evidence to suggest that certain groups may be particularly at risk of being incorrectly rejected for Settled Status.

Specifically, the algorithm used within the Settled Status process does not look at DWP data concerning Child Benefits and/or Child Tax Credits. The Coram Children's Legal Centre has identified that this decision can impact negatively on women.[86] The Coram Children's Legal Centre argues that limiting the databases interrogated by the algorithm in this way, places women at a disadvantage because they are more likely to be in receipt of Child Benefit (it is payable only to the primary parent) and/or Child Tax Credits than men. There seems to be good reason to conclude that this may be the case since in August 2018, 87% of Child Benefit recipients were female and 12% were male.[87] Equally, there is clear statistical evidence released in January 2018 showing that women are more likely to be in receipt of Child Tax Credits, whether they be single parents or in a couple.[88]

If this is the case, then the government would need to be able to justify the system objectively. In the authors' view, the government would find it difficult to justify its use of the Settled Status system if it were to rely on the aims outlined above.[89] It is not possible to currently assess the effectiveness of the system overall in terms of accurately identifying who has five years' of continuous residency. However, it does appear that any justification defence would fail since there are means of achieving the government's aims which are less or non-discriminatory.

---

data/file/799413/EU_Settlement_Scheme_public_beta_testing_phase_report.pdf

86 See https://www.childrenslegalcentre.com/wp-content/uploads/2019/03/EUSS-briefing_Mar2019_FINAL.pdf

87 See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/782370/ChB_18_commentary_pdf.pdf

88 See figure 7.1 at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/677582/cwtc-main-Dec17.pdf

89 Such as reducing the Home Office's reliance on paper documentation, reducing caseworker processing time, reducing fraud and error and improving the customer journey.

Put quite simply, it is difficult to understand why it did not interrogate the Child Tax Credit and/or Child Benefit database in the same way as the other benefits data base. This would, it appears, cure the discriminatory impact of the current system under discussion here whilst presumably improving its accuracy. Whilst it is recognised that one of the aims underpinning the system was the need to quickly address a large number of applications within a short period of time, it is difficult to see why interrogating this extra data base via the algorithm would have caused any significant delay or other difficulty.

Moreover, it is to be expected that utilising more relevant data under the control of the government would lead to more accurate results and a lower need to require additional paper documentation. In other words, it is surely a step which would have improved rather than hindered the system which, if correct, would almost certainly be fatal to any justification defence.[90] In the authors' view, therefore, excluding this relevant data, for such a significant number of people for apparently no good reason, is unjustifiable in light of the disparate impact on women.

Data protection rules and Europe's existing equality laws

One of the challenges to AI is that is cannot simply be analysed within an equality framework. Since "big data" is central to many forms of AI, it is also crucial for Equinet's Members to understand the interplay between data protection rules and Europe's existing equality laws.

European Union Data Protection

The starting point is that Article 8 of the Charter of Fundamental Rights of the European Union (CFREU) enshrines the right to data protection.

> Everyone has the right to the protection of personal data concerning him or her.
>
> Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.

90 It should be noted that there may be technical reasons for the decision not to interrogate this data, which would have to be assessed by a court.

The General Data Protection Regulation (GDPR)[91] together with the Law Enforcement Directive (LED)[92] give more detailed protection of natural persons with regard to the processing of personal data (Article 1). Equinet's Members should also be aware of the related Privacy and Electronic Communications Directive (PECD),[93] which supplements the GDPR and LED, but a discussion of the reach of PECD is outside the scope of this report.

Under the GDPR, data subjects have a right to object to the use of algorithms and ML under Article 21 (1) even if processing would otherwise be lawful, in certain limited circumstances –

> The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims.

Article 6(1)(e) and (f) state as follows –

> Processing shall be lawful only if and to the extent that at least one of the following applies:
>
> ...
>
> (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
>
> (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Equally, under Article 22 of the GDPR, a data subject has the right not to be subject to decisions made in consequence of the pure application of an algo-

91 See https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

92 Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA; see https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0680

93 Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector; see https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32002L0058

rithm (whether or not underpinned by ML) where there are legal consequences for him or her or similarly significant repercussions, including decisions that are discriminatory. Article 22 states

> Everyone has the right to the protection of personal data concerning him or her.
>
> Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
>
> 1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
>
> 2. Paragraph 1 shall not apply if the decision:
>
> (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;
>
> (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
>
> (c) is based on the data subject's explicit consent.

It will be a matter for the courts ultimately to determine whether automated processing produces legal effects, but there appears to be no reason to construe this Article restrictively. Guidance can be sought from the European Data Protection Board (EDPB).[94] However it is also clear that, like Article 21, the right created by Article 22 is limited in many ways as set out in the full text.

The LED[95] covers the protection of natural persons regarding the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences, criminal penalties and the protection of public security: Article 1. It also applies in relation to cross-border processing of personal data for law enforcement purposes. The LED does place limitations on the processing of data which might be relevant to protected characteristics like race.

These two sister provisions, the LED and the GDPR, are intended to complement one another and regulate entirely different spheres. Accordingly, Article 2 (2)(d) of the GDPR expressly "carves out" the matters which fall to be regulated by the LED.

94 See https://edpb.europa.eu/edpb_en

95 See https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0680&from=EN

Using the GDPR to open the "black box"

Importantly, the GDPR specifically refers to the principle of transparency; there is now an important debate in Europe as to the extent to which these principles amight be used to force organisations to disclose the contents of their "black box". The current position of the EDPB is that the GDPR does not go so far as to dictate that algorithms or the basis for ML must be disclosed.[96] It simply considers that -

> The GDPR requires the controller to provide meaningful information about the logic involved, not necessarily a complex explanation of the algorithms used or disclosure of the full algorithm. The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision.

Thus, while the GDPR may be useful in terms of seeking to look inside the "black box", the EDPB does not currently consider it compels complete transparency. So it cannot be said that Article 22 is a tool which will always secure full compliance with equality law.[97] Whether the EDPB guidance conforms to Article 22 has not yet been reviewed by the CJEU; if that were to happen it may be that the CJEU would hold that it should be interpreted in a way which is fully coherent with equality law. It is not however appropriate just to wait and hope that happens. It is recommended that this issue is raised specifically with the EC in response to the White Paper of February 2020.[98]

As some disclosure of the proposed use of AI is necessary to comply fully with Article 22, it is a small step to say these potential uses should be publicly listed in some form. This possibility has been discussed already; for instance it has been suggested that states set up a registry of the uses of AI,[99] and in Malta, a certification scheme is already in place;[100] see further Appendix 2.

Equinet and its Members should point to these proposals in their responses to the EC's White Paper, pointing out the importance of the clear statement by the

96 See https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

97 The extent to which this is sufficient guidance is thus controversial. See for instance the CoE's Report on developments after the adoption of Recommendation (2010)13 on profiling, prepared by its Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data - Convention 108; see https://rm.coe.int/t-pd-2019-07rev-eng-report-profiling/168098d8aa

98 See "On Artificial Intelligence – A European approach to excellence and trust"; see https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

99 See https://algorithmwatch.org/en/story/kees-verhoeven-algorithm-registry/

100 The way in which Malta's certification system works is described at https://www.maltachamber.org.mt/en/malta-first-country-in-the-world-to-launch-ai-certification-programme

AI HLEG that "explainability" is necessary.[101] In considering the way forward the EC will need to ensure that any future legislation will enable the "black box" to be fully opened to the extent necessary for equality rights to be fully secured.

One further point concerning the "black box" problem should be noted here. There is increasing concern that developing world trade rules concerning the digital economy will seek to protect source code and algorithms, in a way which is detrimental to transparency.[102]The EC's White Paper contains a section relating to "International Aspects" [103] but has not directly discussed the tension between on the one hand securing a justifiable protection of trade secrets, and on the other, the necessity to be able to see and understand how AI can cause unequal treatment. Equinet and its Members should point out how important it is that transparency is not undermined in this way.

While it seems unlikely that the EU would permit negotiations on international trade in this field to cause any diminution in the protections found in the GDPR, the EU and those states which negotiate on their own behalf, must ensure more positively that the rules of international trade in the digital market are always consistent with providing all necessary transparency to protect the principle of equality.

Country specific prevention of discriminatory AI and data protection laws

This Report now moves to look at specific AI initiatives within Equinet's Membership. Throughout the period in which research has been undertaken, it has not been possible to identify any examples of countries within Equinet's Membership where AI specific legislation has been enacted to expressly tackle discriminatory systems.[104] Although, there are some legislatures which are considering the issue, such as France[105] and Germany,[106] and others such as the

101 See the discussion of "explainability" above and see https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1

102 See e.g. McCann D., "e-Commerce Free Trade Agreements, Digital Chapters and the impact on Labour: A comparative analysis of treaty texts and their potential practical implications", 2019 New Economics Foundation, published by the International Trade Union Confederation: see https://www.ituc-csi.org/IMG/pdf/digital_chapters_and_the_impact_on_labour_en.pdf

103 See the White Paper, op. cit. at Section H - International Aspects.

104 There were inevitably some countries where additional legislation had been enacted in light of the GDPR and for other data protection reasons, which do regulate automated decision making, but the authors have not been able to identify legislation which is intended to target expressly discriminatory artificial intelligence.

105 See https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

106 See https://www.bundestag.de/en/committees/bodies/study/artificial_intelligence

United Kingdom which are consulting on what steps should be taken to ensure that existing regulatory provisions are properly understood.[107]

Numerous countries have enacted specific domestic legislation to implement the GDPR and LED, with some, like the UK, adding their own interpretation of parts of this framework. Similarly, many countries have bespoke data protection legislation, some of which appears to create a stronger principle of transparency in relation to algorithms than the GDPR. A more detailed account is given in Appendix 2.

There is, however, a fine line between substantive law and procedural provisions and one commentator has noted that eight EU member states (Belgium, the Netherlands, France, Germany, Hungary, Austria, the UK, Ireland) do provide specific exemptions and relevant safeguards pursuant to Article 22(2)(b) GDPR.[108]

A similar recommendation has been made by consultants retained to advise the government of the Netherlands,[109] who propose that if new legislation is put forward it should include a similar obligation[110] –

> Algorithm impact assessment: let public agencies that want to use algorithms perform an impact assessment before it is decided to use them. This tool will help public administration to identify risks, mitigate them and check whether risks remain high. Arrange for authorities that have identified high risks but wish to use the algorithms to request the designated supervisor for a „prior consultation".

The French approach to ensuring that those who are affected by administrative decisions which are based on AI systems are notified of that fact is particularly

---

107 See the Information Commissioner's Office and the Alan Turing Institute's consultation on "Explaining AI decisions guidance"; see https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-and-the-turing-consultation-on-explaining-ai-decisions-guidance/. The Centre for Data Ethics and Innovation is also undertaking a "Review focusing on Bias in Algorithmic Decision-Making." which is due to be published Late Spring or Summer 2020; see https://www.gov.uk/government/publications/responses-to-cdei-call-for-evidence/cdei-bias-review-call-for-evidence-summary-of-responses

108 Malgieri, G., 2019. Automated Decision-Making in the EU Member States: The Right to Explanation and Other'Suitable Safeguards' for Algorithmic Decisions in the EU National Legislations. Computer Law & Security Review., see https://www.sciencedirect.com/science/article/pii/S0267364918303753

109 See Prof. dr. Valerie Frissen, dr. Marlies van Eck Thijs Drouen LLM of Hooghiemstra & Partners, Research Report on Supervising governmental use of algorithms, 2 January 2020, see https://hooghiemstra-en-partners.nl/wp-content/uploads/2020/01/Hooghiemstra-Partners-rapport-Supervising-Governmental-Use-of-Algos.pdf

110 Ibid. at p. 5; see also p. 25.

significant.[111] This does not specifically relate to equality but does go some way to ensuring that there is greater transparency about the way in which an individual is treated.

The work of the German Data Ethics Commission should also be noted; in October 2019 it published an Opinion[112] concluding[113] that –

> Consideration should be given to expanding the scope of anti-discrimination legislation to cover specific situations in which an individual is discriminated against on the basis of automated data analysis or an automated decision-making procedure. In addition, the legislator should take effective steps to prevent discrimination on the basis of group characteristics which do not in themselves qualify as protected characteristics under law, and where the discrimination often does not currently qualify as indirect discrimination on the basis of a protected characteristic.

In parallel with this work, there are many examples of countries in Europe looking at the ethical implications of AI. Academics and pressure groups are also examining these matters. A more detailed account is given in Appendix 2.

Equinet's Membership and legal reform

To summarise, Europe has equality rules and data protection rules which are highly relevant and important to the fight against discriminatory AI. However, there is more work to be done to ensure that in all European territories within Equinet's Membership there is a meaningful legal framework. It follows that the interplay between equality and data protection is an area where Equinet's Members are uniquely placed to act. As explained, whilst data protection might ordinarily not have fallen within the remit of Equality Bodies, the challenges posed by AI mean that they must be ready to engage with data protection rules so as to ensure that the principle of equality is respected. This process must be undertaken alongside a review of the existing equality framework.

111 See the discussion in the note written by Marlies van Eck, The use of algorithms by the government: what's the French administrative law secret?, 3rd January 2020, at https://automatedadministrativedecisionsandthelaw.wordpress.com/2020/01/03/the-use-of-algorithms-by-the-government-whats-the-french-administrative-law-secret/ and see article L 311-3-1 Public relations code and administration, added by article 4 of the Loi pour une Republique numerique, Loi n° 2016-1321 see https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte

112 See https://datenethikkommission.de/wp-content/uploads/191023_DEK_Kurzfassung_en_bf.pdf

113 Ibid. at Recommendation 53.

CHAPTER 4: THE WORK THAT EQUINET'S MEMBERS ARE CURRENTLY UNDERTAKING TO ADDRESS THE DISCRIMINATORY EFFECTS OF AI SYSTEMS

Survey of Equinet's Members

A survey of Equinet's Members was conducted to establish what work they were already undertaking in relation to AI in collaboration with Equinet, an online survey was designed to assess the extent of their activity.

The survey was sent to all of Equinet's Members and was open for response between July to September 2019. Responses from 30 organisations were received, some of whom provided multiple responses. There was also an opportunity to participate in an exchange of ideas between Equinet's Members in Paris on the 26 September 2019 in a meeting to discuss AI, organised by the CoE.

This provided the authors with an opportunity to speak via email and / or interview about AI and its challenges to representatives from: Unia (Belgium), the Defender of the Rights (France), the Federal Anti-Discrimination Agency (Germany), the Human Rights and Equality Commission (Ireland), and the Institute for Human Rights (Netherlands). Additional information arising from survey responses was further provided by the Institute for Human Rights (Netherlands), Unia (Belgium) and the Defender of the Rights (France).

The wealth of information received has allowed the identification of various themes which are developed in this report.

Knowledge of relevant law and connections with other agencies

The Survey results are set out more extensively in Appendix 3. In this chapter attention is drawn to some of the key aspects that emerged.

Despite the prevalence of AI within Europe, only 70% of survey respondents were aware of a debate within their country's legislative body concerning the regulation of algorithms and AI (see Appendix 3, figure 4) and only 60% of respondents were aware of a public debate within their country concerning the potential of AI to discriminate (see Appendix 3, figure 1).

It appears that even where regulation is being considered, a good number of legislative bodies are not considering the equality implications of these new forms of technology. Considering our analysis in Chapter 3 above, in which the Report examines precisely how these technologies can discriminate, this is a worrying omission.

On the other hand, to the extent that Equinet's Members are engaging with the public discourse on AI, they are aware of it involving a range of actors including private companies, individuals (via social media), mainstream media, charities, pressure groups and academics (see Appendix 3, figure 2).

It is noted that the debate is mostly focused on the protected characteristics of gender, race and disability, and, only to a lesser extent, on sexual orientation, pregnancy, religion/belief and age (see Appendix 3, figure 3). This should be a matter for discussion within Equinet. While race, gender and disability are undoubtedly important so are these other characteristics which have received less attention. In particular, it is surprising that there was less engagement with pregnancy and age as protected characteristics. Pregnancy is of course closely aligned with gender issues, but it is a separate characteristic, nonetheless. In a rapidly ageing Europe,[114] it is very surprising that more focus has not been placed on the characteristic of age too.

Interestingly, one of the themes which emerged from the survey and discussions with the Equinet membership is that, with a few exceptions, most organisations did not appear to be fully aware of concrete ways in which AI and algorithms were being used in their countries. It appeared that media coverage had led to a general understanding that AI must be "out there", but Equinet's Members[115] seemed to have had little access to reliable information about what was happening. No doubt this has been exacerbated by the fact that, unusually, direct complaints were rarely received from the public. This is no doubt a consequence of the lack of transparency as to the use of AI and it highlights the urgent need for awareness raising by Equinet's Members among the public so that a virtuous circle is created in which an informed public raise these issues and Equinet's Members address them.

During the research stage for this Report, it has been very difficult to identify public and/or centralised sources of information concerning the use of algorithms, AI and automated decision making (ADM). This will inevitably have contributed to Equinet's Membership having limited insight into the way in which AI systems are being used. The only organisation which was seeking to collate

114 See https://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing

115 For example, news items to the effect that tech companies like Facebook allowing organisations to control who can see advertisements on the basis of protected characteristics like race (https://www.technologyreview.com/s/613274/facebook-algorithm-discriminates-ai-bias/?utm_campaign=site_visitor.unpaid.engagement&utm_source=twitter&utm_medium=social_share&utm_content=2019-09-30) and that companies using algorithms to monitor employees whilst at work (https://theconversation.com/worker-protection-laws-arent-ready-for-an-automated-future-119051?utm_source=twitter&utm_medium=twitterbutton).

meaningful detail about the use of these forms of technology within Europe was the website of AlgorithmWatch.[116] The information gathered is set out in Appendix 1, but it must be acknowledged that this is likely to provide more of a "snapshot" of what is happening in Europe rather than a comprehensive narrative.

A further theme identified from the survey, and which was repeated in discussions, was a common concern within Equinet's Members that they lacked the technical expertise to be able to address discrimination caused by algorithms and artificial intelligence. Equally, there was a very real concern that they lacked the resources to be able to adequately address the discrimination which new forms of technology might be creating.

The Equinet survey also showed that many of members were unaware of work in this area already being undertaken by the European Union (EU) and the Council of Europe (CoE) which is outlined above and in greater detail in Appendix 2. 30% of the respondents, did not know about the EU's projects (see Appendix 3, figure 9) and just under 20% did not know about the CoE's initiatives (see Appendix 3, figure 10). A programme led by the EU and CoE is needed to ensure that Equinet members are aware of these initiatives so that they can effectively challenge discrimination. Equinet can also assist in this task.

Another very significant finding was that, despite the obvious link between AI and other areas such as data protection, only around 50% of survey respondents were actively considering working with other organisations within their country to tackle discrimination (see Appendix 3, figure 8).

Specifically, of the respondents, 12 explained that they did not know anything about their country's steps to comply with these provisions within the LED (see above) and 7 equally had no knowledge of the equivalent provisions with the GDPR (see above). It is not clear why there was so little connection being made. It may be linked to a lack of awareness of the detail of European data protection legislation and that it regulates certain forms of ADM, as this Report has noted above. Again, Equinet has an important role in encouraging and facilitating its members to be fully aware of the way in which equality law and data protection law operate together to control AI systems. Regulation in relation to AI systems cannot be undertaken in silos. It would be quite wrong for each regulator to undertake its designated tasks on a narrow basis and to ignore the connections with the related concerns of other regulators with other functions.

116 See https://algorithmwatch.org/en/

Initiatives by Equinet's Members

Strategic plans

Despite the slightly negative picture the survey presented, some of Equinet's Members have started to formulate structured "campaigns" to tackle discrimination and AI -

- The Defender of Rights in France has started to work on a formal project to tackle algorithmic discrimination.[117] It is considering working with the data protection body in France and academic or other organisations with technical expertise.
- In Germany, the Federal Anti-Discrimination Agency recently funded a study on *"Discrimination risks concerning the use of algorithms"* which was released in September 2019.[118] The Equality Body plans to use this study as a springboard from which to lobby the government on change.
- Information was shared on a confidential basis by one member that it is undertaking research into algorithmic discrimination, but the authors were not able to obtain any further information about this project.[119]

In addition, over 70% of the respondents answered that they were *considering* the impact of algorithms and AI from an equality perspective (see Appendix 3, figure 5) and over 60% were *considering* acting or had started taking action (see Appendix 3, figure 6). This intention to act is important since 70% of the respondents were not currently defining best practice in this area (see Appendix 3, figure 7). This result shows that action by Equinet would be welcomed by its members.

There is plainly a mood among members to do more in this area and so any expertise that Equinet can bring is likely to be of real assistance. There is therefore more than just a mandate for Equinet to work further on these issues but a potentially very receptive audience for its work among its membership.

Decisions by equality bodies

Some important examples of Equinet's Members using their powers to issue decisions in relation to algorithms were identified -

117 Equinet survey research.

118 See https://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/Expertisen/Studie_Diskriminierungsrisiken_durch_Verwendung_von_Algorithmen.pdf?__blob=publicationFile&v=4 (with summary in English).

119 Equinet survey research.

- The Defender of Rights in France has used its powers to publish opinions and decisions outlining its concerns into the use of ADM. There have been decisions into the use of algorithms in relation to the scoring practices in banks and insurance companies[120], a decision in April 2019 setting out criticisms of Parcoursup[121] and an opinion in October 2018 which outlined concerns about the use of algorithms in the justice system[122].
- An Equality Ombudsman investigated a matter in which a national state-owned bank had used an algorithm to calculate the credit risk for an individual aged over 60 years old in 2018.[123] After the investigation was completed, the bank changed its rules.

However, the impression given by Equinet's Members is that they have taken this action because matters had been escalated to them on an *ad hoc* basis. This work is commendable; however, a much more strategic approach is necessary to ensure that both commerce and public bodies are aware that they must take steps to ensure that the use of AI systems is not discriminatory.

Sector specific projects

Some of Equinet's Members have collaborated directly with governmental bodies and private companies in order to discuss AI and discrimination in sectors which were considered particularly sensitive. For example, the Non-Discrimination Ombudsman in Finland has liaised directly with governmental organisations responsible for examining AI and ethics in Finland, including the Ministry of Justice, to ensure that the principle of non-discrimination is fully considered. [124]

It has also liaised directly with representatives of the banks, insurance companies and finance to ensure that discrimination does not occur in relation to ADM.[125] The Office of the Equal Opportunity Ombudsman in Lithuania has also met with financial institutions including banks and insurance providers to discuss algorithmic discrimination and it has made recommendations for insurance companies in relation to selling travel insurance to people aged over 65.[126]

Well-informed and timely interventions by Equinet's Members in the discussions within government are to be welcomed if they can help steer the discourse

120 See https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=12969

121 See https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=18803

122 See https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=18058

123 Information provided at the Paris meeting in September 2019.

124 Equinet survey research.

125 Equinet survey research.

126 Equinet survey research.

and administrative actions initiated by those bodies in a positive way. The key to success though is that the Equality Body is well–informed and can therefore make an expert and well–respected contribution.

Having outlined the current state of the public discourse in Europe, the ways in which the debate can be further moulded and advanced will be considered.

CHAPTER 5: PROGRAMME OF ACTION FOR EQUINET'S MEMBERS AND THEIR STATES

Equinet's Members can undoubtedly build on the work that is already being done and further positively shape the debate concerning AI and equality.[127] Moreover, Equality Bodies are uniquely placed to fulfil this task. Whilst challenging new technology requires a multi-disciplinary approach, there is a human rights and equality dimension to very many of the new forms of AI which are being used across Europe. Equinet's Members are uniquely placed to leverage their existing expertise and knowledge to meet the challenges posed by AI.

In this chapter, this Report sets out how a generic approach might be taken by Equinet's Members and posits a range of possible interventions to enhance their ability to meet the new challenges they face. It also addresses the role of states in ensuring that Equinet's Members can carry out the work identified as necessary by (i) this Report, (ii) Equinet's Members carrying out the work recommended by this Report, and (iii) in order to ensure compliance with best practice as it is developed by the work of the CoE and the European Union.

Resourcing

It is absolutely crucial that Equinet's bodies are adequately resourced to address the important interplay between AI systems and the principle of equality and non-discrimination. Equinet's Members face a significant and important task because of the unique challenges which AI creates, including its cross-sectional nature and the broad range of circumstances within which it is being utilised.

In the most part Equinet's Members will rely on funds and resources granted to them by their respective states. It is therefore critical that such states understand the issues that confront them as AI develops, and work with Equinet's Members to ensure that they understand the resources that these Equality Bodies need to carry out their work. At a time when all states are under financial pressure, states must not under-estimate the importance of this. The rapidly increasing use of AI is having effects on almost every aspect of society. States that do not respond swiftly to the need to resource Equinet's Members will inevitably find that they are at a very great disadvantage in securing that AI is used properly to the benefit of their societies.

127 An overview of the powers available to each Equality Body is helpfully summarised at http://equineteurope.org/what-are-equality-bodies/european-directory-of-equality-bodies/

Resources must be found to bring the level of expertise within Equinet's Members up to the necessary standard to understand what is occurring and to be able to develop the work programmes that will be needed. The resources may be broadly divided into two categories. First resources will be needed to recruit and develop the basic level of experience needed within each Equinet member. This will involve defining the person specification and job description for the team member or members that will be needed. It will then involve the process of finding or training the persons to fit these criteria. The second level of resource will involve the ongoing maintenance of this team and the costs of its deployment within each state. The assessment of these costs will depend on the way in which each Equinet member responds to the points made in this report.

Mandate

It is also quite clear that each of Equinet's Members will need to review the legislation or administrative provision under which they exercise their functions. These country specific mandates will of course differ and some of them may be sufficiently wide to take on all the new functions that this Report has identified. However this is unlikely to be correct for all countries and so where appropriate and necessary each Equinet member should seek a strengthening of its powers to ensure that it can meet the challenges that the development of these new technologies pose for equality and non-discrimination. States must respond to these requests to ensure that Equinet's Members can meet the challenges that this Report has identified.

In particular, as outlined further below, it will be crucial for Equinet's Members to have powers to conduct investigations and to impose sanctions for non-compliance.

Mapping the territory

The first urgent step is to understand better what is happening in the countries where each of Equinet's Members operate. As this Report has noted, generally speaking, members are struggling to understand the full extent to which AI is being used in their countries. This is entirely understandable, since there is limited information which is publicly available concerning the use of algorithms especially in the private sector. However, this must be addressed as a matter of some urgency. Inevitably, a lack of knowledge will mean that it is difficult for organisations, such as Equinet's Members, to engage in a public discourse in which the risks of AI and possible discrimination can be meaningfully highlighted. It will always be more effective to campaign for change when conc-

rete examples of harm or potential harm can be highlighted rather than simply vague, theoretical risks.

In the next paragraphs recommendations are made as to how members might obtain better information about the situation in their area of responsibility.

Public Inquiries

Equinet's Members should consider launching public inquiries into the use of AI within their own countries. An inquiry of this type could involve interviews or public hearings with public sector and private sector actors to uncover the true extent to which AI is being used.

Domestic desktop reviews

An alternative, which might be more cost-efficient though perhaps not so comprehensive, would be to commission a research project which could undertake a "desktop" review of the use of algorithms and AI systems. This could be carried out in conjunction with an academic institute, charity or another regulator.[128]

Europe-wide thematic reviews

A further idea is that Equinet's Members might wish to co-ordinate their enquiries / research projects across Europe in a wider thematic review. As identified in Chapter 2 above, there are common themes as to the way in which AI is being deployed across Europe. It may be advisable for Equinet's Members to limit their enquiries to sensitive sectors (e.g. social advantages, education or financial services). Equinet's Members could take this approach one step further and essentially "divide up" different sectors between them to avoid replicating work whilst ensuring that a complete picture is obtained.

Legal "gap analysis"

Once Equinet's Members can highlight concrete ways in which AI is being used with the potential to breach the principle of non-discrimination, it would be advisable to conduct a legal "gap analysis". The purpose of this analysis would be to understand the way in which AI systems should be analysed within a country's

128 For instance, in the UK, the statutory Advisory Conciliation and Arbitration Service, tasked with resolving workplace disputes, published a Report entitled "My boss the algorithm: an ethical look at algorithms in the workplace" produced in conjunction with Patrick Briône of the Involvement and Participation Association; see https://www.acas.org.uk/my-boss-the-algorithm-an-ethical-look-at-algorithms-in-the-workplace/html#executive-summary

individual legal system. The analysis should help clarify what can be addressed within existing laws and where there is a need for legal reform to ensure that AI is properly regulated.

This "gap analysis" could then be used as a springboard from which Equinet's Members could campaign for changes to existing legal frameworks in order to "fill" any gaps in legal protection or alternatively enhance existing legal frameworks.

The kinds of policy, legislative and regulatory issues, that such a "gap analysis" might address, include the need for legal reform in respect of the following -

- Whether the principle of non-discrimination as enacted in each state is insufficiently broad (see Chapter 3). In particular it is possible that the provision of goods, facilities and services is not protected from discrimination in relation to all of the protected characteristics in the same way as work and employment. These reviews should look specifically at the delivery of social advantages and other public goods.
- Whether it is clear who will be liable for any discriminatory AI. There can be confusion between the liability of an end user of AI, the company that supplied the technology or the body or organisation that supplied any relevant data sets. Any or all of these could be made liable and it is of course essential that at least one such body is made comprehensibly liable for any discriminatory effects. If this is not clear then legislation will be need to be amended to make it clear that, at the very least, the party who places a product in the market is legally liable for any discrimination in accordance with the EC's recommendations (Appendix 2).
- Whether there is a sufficiently developed principle of transparency such that it is impossible to identify whether an AI system is non-discriminatory (see Chapter 3 and Appendix 2). There is a near universal acceptance that the difficulty in providing full transparency cannot justify discriminatory outcomes. The problem is to secure that such transparency obligations are effective and adequate. Equinet's Members might wish to focus on ensuring that the principle in *Danfoss*,[129] namely that there is an inference of discrimination where a system lacks transparency, is enshrined clearly in domestic legislation in the context of artificial intelligence. In this respect Equinet's Members might wish to champion the EC's proposal for a reversal in the burden of proof (Appendix 2, para 21).
- Whether there are sufficient appropriate procedural safeguards so as to limit the use of discriminatory technology. For example, a requirement to produce Algorithmic Impact Assessments or Audits, similar to Data

129 Case C- 109/88, Handels- og Kontorfunktionaerernes Forbund i Danmark v Dansk Arbejdsgiverforening Ex p. Danfoss A/S ECLI:EU:C:1989:383.

Protection Impact Assessments, which demonstrate that the potential for the technology to discriminate has been assessed and minimised (see Appendix 2).

- Whether a public register or certification might be useful. Many countries are experimenting with new legal approaches towards the regulation of algorithms such as the creation of certification schemes (see Appendix 2). A public enquiry supported by a "gap analysis" might also give rise to similarly creative new ideas to improve the regulation of AI.
- Whether a fresh binding human rights protocol would be useful. For instance the Netherlands Institute of Human Rights has stated that if such a protocol were created then it would be able to lobby the Dutch government to ratify it.[130] Equinet might consider developing a draft protocol based on common agreement of the problems to be faced and the deficits in protection that its members have found.
- Whether the interplay between data and equality requires an entirely new approach to be adopted towards tackling AI issues. A specialist body of the EC has commented that "National AI strategies have so far paid little attention to the challenges AI poses to the [intellectual Property Rights] legal framework."[131] Others have suggested that a specialist courts might be necessary to review and monitor AI, possibly with rules that limit the extent to which sensitive commercial data could be shared publicly.[132] Equinet's Members will need to consider the implications of this kind of problem and might for instance propose that they take on the adjudication of this kind of problem issuing recommendations and other sanctions.
- The EU, and those states that are not within the EU which will be negotiating their own international trade rules concerning the digital economy, will also need to assess whether such rules interfere with the protection of equality and the elimination of discrimination by AI systems.

In parallel with this "gap analysis" Equinet's Members need to consider how adequate guidance can be communicated to the public and how awareness of

the issues can be raised within their states. Equinet's Members might propose binding legal guidance and Codes of Practice.

130 Equinet survey research.

131 See the Report of the European Commission's Science and Knowledge Service, Joint Research Centre: "Artificial Intelligence - A European Perspective", 2018 at [7.3]; see https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf

132 Lord Sales' speech Algorithms, Artificial Intelligence and the Law was given at the Sir Henry Brooke Lecture for BAILII, London and is available here: https://www.supremecourt.uk/docs/speech-191112.pdf

Further work on developing ethical principles

There is a wide range of ethical principles currently being developed by numerous actors.[133] Equinet's Members should also play a leading role in developing the national understanding of the ethical principles that must apply to the proper use of AI systems if they are to conform to the principles of equality and non-discrimination. This will be particularly important since, as explained in Chapter 3, as the law develops, organisations who utilise AI will only be able to justify it in so far as it is compliant with nationally accepted ethical standards. Some of the opportunities for Equinet's Members to engage in this process are discussed in Appendix 2.

Non-legally binding guides

Equinet's Members could play a valuable role in producing guides and other communications (e.g. advertising, social media campaigns, websites, seminars, tool kits) which explain to the public, companies, state actors, data scientists, developers and legal professionals, how discriminatory technology is prohibited and should be analysed. There is always a need for guides to the effective implementation of equality and non-discrimination principles.

Test or strategic litigation

Equinet's Members should also consider bringing, supporting, or funding test or strategic litigation that challenges discriminatory technology. The aims of such litigation could be to clarify the law and also to publicise the potential for the principle of non-discrimination to be infringed by emerging forms of technology. Current litigation in Europe which is challenging AI is highlighted later (see Appendix 2).

Collaboration with other regulators

It is crucial that Equinet's Members liaise and collaborate closely with other regulators across the spectrum of possible equality problems, so that a "joined up" approach is adopted. In light of the sheer breadth of issues identified in Chapter 2 and further explored in Appendix 1, the author's foresee that Equinet's Members will need to work, at the very least, with regulators concerned with –

- data protection,
- consumer protection,

133 Algorithm Watch has produced a detailed "inventory" of current AI ethical principles. It is available here: https://inventory.algorithmwatch.org/ and demonstrates the sheer array of ideas in this area.

- employment rights,
- health care and
- financial services.

Each of these will have their own designated regulatory powers, but they will also be subject to equality laws and will need to ensure that their specific obligations do not enable or permit discrimination by AI systems.

Training the public and the coders on equality

Equinet's Members could play a key role in education and training. The Equality Body in the Netherlands has developed a training programme which it delivers alongside an external provider to employers to highlight how discrimination can infect recruitment processes. The feedback received after the training was extremely positive. There is the potential for Equinet's Members to replicate this model in relation to AI and algorithmic discrimination targeting both local authorities, governmental bodies and private companies. Finland's Equality Body also suggested that creating a training or certification programme for organisations that used AI would be a positive step. A similar scheme was introduced in Finland in relation to housing which meant that as part of the relevant qualification for estate agents, it is obligatory to complete a section on discrimination. This change was introduced as a result of the work of the Non-Discrimination Ombudsman.[134]

In addition to informing and training the public, Equinet's members should also aim at training future coders through contributions to the faculties of national universities and other academic institutions and through cooperating with standardization initiatives such as those undertaken by the Institute of Electrical and Electronics Engineers (IEEE).[135]

Equality by design

A related point is that Equinet's Members should consider producing, or working alongside regulators, to produce practical guidelines aimed at public actors, businesses and service providers who are creating or deploying AI so as to ensure that systems are designed so as to be free from discrimination. For instance, in the UK, the government has sought to embed ethical design into its procurement processes (see Appendix 1). There is great potential for Equinet's

134 Equinet survey research.

135 See for instance IEEE's "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition (EAD1e)"; see https://ethicsinaction.ieee.org/

Members to play a leading role in ensuring that the creators of AI are designing systems in a way which are non-discriminatory.

Data scientists and other experts

One issue highlighted during the research for this report, is that many of Equinet's Members feel unable to engage effectively with AI because of a lack of understanding the technology itself. There is also a concern that Equinet's Members are not sufficiently resourced to be able to "buy in" relevant expertise. Adequate resources are indeed critical, but they may not always be as expensive as may be feared. For instance, the UK's Information Commissioner's Office recently appointed its first Postdoctoral Research Fellow in Artificial Intelligence to support its work.[136] This fusion of academia and regulatory work may prove to be a useful model.

Developing understanding of AI and Equality

Equinet's Members should be active contributors to the production of relevant AI-knowledge through their involvement in national and European expert bodies working on strategies and legislation for AI. They should work with universities and similar academic institutions and expert groups. By partnering in this way with such expert groups they will enhance the value of the discussions those organisations have with the EU and the CoE. This could have a double benefit, raising awareness for anti-discrimination in AI and enhancing the AI-related expertise of Equinet members.

136 See https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/11/information-commissioner-s-office-appoints-in-house-expert-to-research-and-investigate-the-impact-of-artificial-intelligence-on-data-privacy/

CHAPTER 6: CHECKLIST

A checklist has been prepared for Equinet so that its members can assess whether AI systems and other similar technologies, including those that use ADM systems, comply with equality and non-discrimination rules.

Direct discrimination		
Q1	Does the artificial intelligence system treat people differently because of a protected characteristic? This could be for various different reasons, such as direct bias in the code used, or in the data set to which it is applied, or because machine learning has treated a particular personal characteristic as a proxy for gender, race ethnicity etc.	Direct discrimination
Indirect discrimination		
Q2a	Does the artificial intelligence system consist of an algorithm and / or is it trained on a data set that places certain protected groups at a disadvantage?	If so, there is *prima facie* indirect discrimination.
Q2b	If so, can the body using the artificial intelligence system point to a legitimate aim to justify the use of the algorithm and / or data set?	Assess the extent to which there is a defence to *prima facie* indirect discrimination

Q2c	If so, is the artificial intelligence system capable of achieving the aim?	Assess the extent to which there is a defence to *prima facie* indirect discrimination
Q2d	If so, is the artificial intelligence system proportionate? Is there a non- or less discriminatory means by which the aim could be achieved? Does the system comply with relevant ethical standards such as transparency and explainability?	Assess the extent to which there is a defence to *prima facie* indirect discrimination
Transparency		
Q3	What information is available to assess the answers to Q1 to Q2 above? Has any audit been carried out? Are there national or pan-European laws such as the GDPR which will allow my Equality Body to understand more about what is going on? Is the system so lacking in transparency that the courts are entitled to infer *prima facie* discrimination?	Consider the evidential burden

Other types of equality claim		
Q4	Are there other equality type claims that could be brought, such as, harassment?	Assess further legal breaches
Data protection		
Q5	Are any data protection provisions engaged here at national level or European level such as the prohibitions in Articles 21 and 22 in the GDPR?	Assess whether there has been a breach of data protection principles
Liability issues		
Q6	Who is the correct party against which any complaint should be made? This might be the developer of the artificial intelligence system or the provider or the party who implements it? The answer will depend on national legislation and the specific facts. It may be that there is more than one party who should be made liable.	Identify the right defendant
Cross over jurisdictions		
Q7	Are there other regulators which might have powers in this area – such as Data or Financial Regulators?	Assess whether other regulatory regimes may apply

Q8a	Are other areas of the law engaged by the artificial intelligence system?	Assess other legal breaches
Q8b	Does public law apply? Since artificial intelligence is being used extensively in the public sector, it is important to be aware that Equinet's Members may be able to challenge the use of algorithms and other technology on the basis that it offends basic principles of public law.[137]	Assess other legal breaches
Q8c	Does competition law apply? There is a growing sense that big data and artificial intelligence may create competition issues. [138] There is already joint action proposed between French and German competition authorities[139] and there is no reason why this should not be undertaken by other such authorities with inputs from Equinet's Members.	Assessing other legal breaches

137 For example, https://www.theguardian.com/uk-news/2019/oct/29/ai-system-for-granting-uk-visas-is-biased-rights-groups-claim & https://adminlawblog.org/2019/12/05/jack-maxwell-and-joe-tomlinson-algorithms-artificial-intelligence-and-the-law-public-law-reflections-on-lord-sales-sir-henry-brooke-lecture/

138 For example, https://algorithmwatch.org/en/story/competition-authorities-ready-for-price-fixing-algorithms/

139 See the Executive summary produced by the Autorité de la Concurrence and the Bundeskartellamt of their joint policy on AI and competition at https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Berichte/Algorithms_and_Competition_summary.pdf?__blob=publicationFile&v=4

Q8d	Have privacy laws been infringed? The prevalence of artificial intelligence may also create privacy issues in that data can be processed by algorithms in order to gather or infer sensitive issues about individuals.[140]	Assess other legal breaches
Q8e	Has employment law been infringed? Artificial intelligence is being used by some employer to monitor their work-force which could give to breaches of national employment law.[141]	Assess other legal breaches
Q8f	Have any consumer laws been breached?	Assess other legal breaches
Q8g	Have any product liability laws been breached?	Assess other legal breaches
Q8h	Have any other fundamental rights been breached?	Assess other legal breaches

140 For example, https://iapp.org/news/a/why-artificial-intelligence-may-be-the-next-big-privacy-trend/

141 See for example, https://www.theguardian.com/technology/2019/apr/07/uk-businesses-using-artifical-intelligence-to-monitor-staff-activity. There is also a concern that there is a growing sector of society, predominantly, gig economy workers who are being essentially managed remotely by algorithm: see https://www.doteveryone.org.uk/2019/10/insights-gig-economy-research/

Identification of the liable party		
Q9a	Who has designed or created the AI system?	Dependent on the course of action, identify the correct defendant
Q9b	Who has created the data set that is used within the AI system?	Dependent on the course of action, identify the correct defendant
Q9c	Who has placed the AI system on the market?	Dependent on the course of action, identify the correct defendant
Q9d	Who is using the AI system in such a way as to give rise to harm?	Dependent on the course of action, identify the correct defendant

APPENDIX 1: USE OF ARTIFICIAL INTELLIGENCE ACROSS EUROPE

It is now impossible to state the full extent to which AI systems are being deployed, since they increase daily. However, this Appendix sets out some of the ways in which AI is being used in a manner which is relevant to Equinet's Membership. This Appendix is not intended to be, and does not give, a comprehensive analysis of each way in which AI, ML and ADM are being utilised. The purpose is only to give a sense of the breadth of the use of this technology and the myriad ways in which it is being deployed.

It will be evident from these examples that the uses identified in one country might just as easily be deployed in any of the other countries in the future.

Austria

Recent media attention has focused on an algorithm deployed by the Austrian Public Employment Service (AMS).[142] According to its website, it is Austria's leading provider of labour-market related services. It matches candidates with job openings and assists job seekers. It deploys an algorithm which automatically assigns a score to each job seeker which then places them in a group: A (people who will likely find employment within a short time), B (people who might benefit from retraining) and C (people who are considered unemployable).[143] This assignment mirrors the discrimination faced by different groups in the Austrian labour market. As a result, people in Group C will receive less assistance from AMS and will be allocated less resources than group B, while they still may be discharged to other institutions. It is said that a human can override the algorithm's assessment. One document (which is not available in English) shows that "*... women are given a negative weight, as are disabled people and people over 30. Women with children are also negatively weighted but, remarkably, men with children are not ...*".[144] The justification for this system is increased efficiency.

Belgium

According to newspaper reports, the VDBA, which is the Flemish employment service, utilises algorithms to monitor the activities of job seekers.[145] Specifically,

142 See https://www.ams.at/organisation/public-employment-service-austria/about-ams

143 See https://algorithmwatch.org/en/story/austrias-employment-agency-ams-rolls-out-discriminatory-algorithm/

144 See http://www.forschungsnetzwerk.at/downloadpub/arbeitsmarktchancen_methode_%20dokumentation.pdf

145 See https://www.nieuwsblad.be/cnt/dmf20170903_03051686

where a job seeker is deemed insufficiently active in terms of looking at vacancies online, they are invited to a meeting with the possibility of a sanction. Resources are also apparently being targeted in certain areas of Belgium using predictive policing.[146]

Denmark

According to AlgorithmWatch, ADM is used in a wide range of scenarios within Denmark such as credit scoring, car insurance, assessing the risk of elderly people requiring additional care, collating employee documentation and breast cancer detection.[147]

Estonia

The CoE's Human Rights Commissioner has stated that algorithms are being widely used within Estonia. Here is an excerpt from a document presented at the High-Level Conference in Helsinki in February 2019, *"Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the rule of law"*[148] –

> In my Report following my visit to Estonia in June last year, for example, I looked at how older persons and their human rights are affected by the use of artificial intelligence and robots in social and care services. NGOs alerted me about difficulties linked to the use of automated decision-making in social benefits services. Following a reform of the work ability support system, machines and algorithms were used to automatically re-evaluate incapacity levels. Reportedly, the incomplete data in the e-health platform, coupled with a lack of in-person interviews, resulted in loss of social benefits for certain persons with disabilities and older persons with disabilities.

Finland

Finland uses AI within "Kela" which is a system that it uses to administer benefits. This is the way in which the system is described by AlgorithmWatch[149] -

> Ongoing and potential AI developments include chatbots for customer service, automated benefit processing, detection (or prevention) of fraud or misunderstanding, and customer data analytics.

---

146 See https://www.standaard.be/cnt/dmf20160517_02292901 & https://algorithmwatch.org/en/automating-society-belgium/

147 See https://algorithmwatch.org/en/automating-society-denmark/

148 See https://rm.coe.int/hlc-helsinki-feb-2019-commhr-intervention-final/16809331b8

149 See https://algorithmwatch.org/en/automating-society-finland/

Equally, the state has used AI to identify risk factors which would indicate that a child might need welfare services in the future including child and youth psychiatry services.[150]

According to the website of the company who developed the AI system, it analysed *"a huge data mass that consisted of client relationship data of the entire population of Espoo for the years 2002–2016 and covered approximately 520,000 people and more than 37 million customer contacts"* which led to the identification of approximately 280 factors that could anticipate the need for child welfare services.[151]

AI is also being used within the recruitment process in Finland to analyse applicants' digital *"footprints"*. This is how the system has been described[152] -

> The Finnish start-up company DigitalMinds is building a 'third-generation' assessment technology for employee recruitment. Key clients (currently between 10 and 20 Finnish companies) are large corporations and private companies with high volumes of job applicants. Personality assessment technologies have been used since the 1940s in job recruitment. At first, these came in the form of paper personality tests that were filled in by prospective job candidates to assess their personality traits. Since the 1990s, such tests have been done in online environments. With their new service, DigitalMinds aims to eliminate the human participation in the process, in order to make the personality assessment process 'faster' and 'more reliable', according to the company. Since 2017 it has used public interfaces of social media (Twitter and Facebook) and email (Gmail and Microsoft Office 365) to analyse the entire corpus of an individuals' online presence. This results in a personality assessment that a prospective employer can use to assess a prospective employee. Measures that are tracked include how active individuals are online and how they react to posts/emails. Such techniques are sometimes complemented with automated video analysis to analyse personality in verbal communication.

The Non-Discrimination Ombudsman in Finland has stated that algorithms are used within the financial services industry in order to credit score individuals. [153]

France

Algorithms are being used extensively in France. The Defender of Rights in France stated that there is a public debate over the use of algorithms in recruitment, the justice system in order to facilitate mediations and health care.[154] In

150 See https://algorithmwatch.org/en/automating-society-finland/

151 See https://www.tieto.com/en/success-stories/2018/the-city-of-espoo-a-unique-experiment/

152 See https://algorithmwatch.org/en/automating-society-finland/

153 Equinet survey response.

154 Equinet survey results and https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=18058 in relation to the justice system reforms.

particular, the Defender of Rights in France has been adjudicating on discrimination caused by statistical assessments within the financial services industry for some years (although it does not appear that sophisticated ML algorithms were being deployed).[155]

Another area of specific concern relates to the education sector and "Parcoursup" which is an algorithmic platform introduced by the French government to select students and assign them to undergraduate courses in an equitable way. Parcoursup uses school records data in order to make a decision which includes the student's place of residency.[156] The Defender of Rights is concerned because Parcoursup moderates students' grades in light of how prestigious their high school is perceived to be and it is felt that disability is inadequately addressed within the algorithm.[157] The French Constitutional Court recently ruled that universities should specify how algorithms have been used to select candidates.[158] It held that once a candidate had been refused admission, he or she may obtain information about the educational reasons for the decision made about them, including information about the criteria used by the algorithms.

FRT has also been used on an experimental basis at two Lycées in the City of Nice. This use of the FRT (reconnaissance faciale) was reviewed by the Commission nationale de l'informatique et des libertés (the French Data Protection Commission) (CNIL), which in October 2019 gave its Opinion, holding that the FRT system, which had the sole aim of making access more fluid and secure for pupils, most of whom were minors, was neither necessary nor proportionate to achieve these goals.[159] In a key part of CNIL's ruling it concluded[160] –

> ...facial recognition devices are particularly intrusive and present major risks of invasion of the privacy and individual freedoms of the persons concerned. They are also likely to create a feeling of reinforced surveillance. These risks are increased when facial recognition devices are applied to minors, who are subject to special protection in national and European texts.

155 See for example https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=12969

156 See https://juridique.defenseurdesdroits.fr/doc_num.php?explnum_id=18803

157 Equinet survey results.

158 Conseil constitutionnel, Décision n°2020-834QPC du 3 avril 2020. See https://www.conseil-constitutionnel.fr/sites/default/files/as/root/bank_mm/decisions/2020834qpc/2020834qpc.pdf An official English translation is not currently available but see the unofficial translation on www.ai-lawhub.com at https://ai-lawhub.com/parcoursup-decision-no-2020-834-qpc/

159 See https://www.cnil.fr/fr/experimentation-de-la-reconnaissance-faciale-dans-deux-lycees-la-cnil-precise-sa-position

160 Ibid.

> ...the Commission recalls that strict vigilance is required in view of the damage that could result from possible security incidents on such biometric data. In this context, and in the presence of less intrusive alternative means, such as using badges as a means of control, the use of a facial recognition device to control access to a school appears disproportionate.
>
> Such a device cannot therefore be legally implemented and it is now up to the region and the high schools concerned, responsible for the envisaged device, to draw the consequences.

It had also been announced that France would start to use an FRT system called "Alicem" in order to create a digital identification system by which its citizens could access government online services.[161] This has proved controversial leading to an announcement in October that the French government would be reviewing the use of FRT.[162]

Algorithms are also being used to combat tax fraud in France. The algorithms employed can analyse various data including information openly shared on social media by individuals. The French Constitutional Court has ruled constitutional this use of algorithms for the purpose of fighting tax evasion, at the condition that the data used shall not reveal any forbidden information, such as race, gender, sexual orientation, political or religious beliefs, genetic and biometric information.[163]

Germany

Companies in Germany have developed "affective computing" systems, whereby personality traits are identified using algorithm by analysing a voice sample for the purposes of human resource management.[164] There has also been a debate about the use of algorithms in predictive policing.[165]

AI is also being used to monitor and plan electricity consumption.[166] In one interesting project, AlgorithmWatch and Open Knowledge Foundation Germany initiated a project called OpenSCHUFA, supported by crowd funding, in which people were asked to donate their credit scores which were then analysed. A

161 See https://www.bloomberg.com/news/articles/2019-10-03/french-liberte-tested-by-nation-wide-facial-recognition-id-plan. A link to the legislation is available here: https://www.legi-france.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000038475477&categorieLien=id

162 See https://www.france24.com/en/20191015-concerns-over-technology-ethics-as-french-politicians-embrace-facial-recognition-3

163 Conseil constitutionnel, Décision n° 2019-796 DC du 27 décembre 2019; see https://www.conseil-constitutionnel.fr/decision/2019/2019796DC.htm

164 See https://algorithmwatch.org/en/story/speech-analysis-hr/

165 Equinet survey results.

166 See https://algorithmwatch.org/en/automating-society-germany/

variety of anomalies were discovered although it does not appear that research was undertaken to link these anomalies to protected characteristics.[167]

Italy

AI has been used in Italy to assign teachers to different regions (and was abandoned following uproar), to help determine the best treatment options for patients, to predict the risk of hospitalisation, by the police as part of FRT, to identify tax evasion and to predict crime.[168]

The Netherlands

Algorithms are being used in commercial settings within the Netherlands in order to tailor products and services. This is an excerpt from AlgorithmWatch[169] -

> ADM in The Netherlands has also found its way into journalism. Several news outlets have implemented, or are in the process of implementing, 'recommender systems'. These systems semi-automatically decide which articles are shown to each individual visitor or subscriber to a news website. Among these outlets are RTL Nieuws, Het Financieel Dagblad, NU.nl and the Dutch Broadcast Foundation (NOS). Most notable among these is a kiosk-like online platform called Blendle that enables users to read articles from multiple newspapers and magazines on a pay-per-view basis. It recently introduced a subscription model that provides subscribers with twenty tailored articles per day. Apart from a few articles that are hand-picked by editors, the selection of these articles is mainly algorithm-based and dependent on a variety of data points (e.g. what articles a user has previously clicked on).

A controversial risk profiling system is being deployed in the Netherlands called System Riscico Indicatie or SyRI by the Department of Social Affairs and Employment with the intention of identifying individuals who are at a high risk of committing fraud in relation to social security, employment and taxes.[170] According to sources, SyRI analyses a wealth of governmental data including identity, education, income and employment.[171] If an individual is deemed to be at a high risk of fraud, then further investigations will be conducted. According

167 See https://algorithmwatch.org/en/automating-society-germany/ & https://algorithmwatch.org/en/schufa-a-black-box-openschufa-results-published/

168 See https://algorithmwatch.org/en/automating-society-italy/

169 See https://algorithmwatch.org/en/automating-society-netherlands/. Please note that the original source material is behind a paywall.

170 The court case challenging this technology was heard in October 2019 and at the time of writing, no judgment had been delivered.

171 See https://pilpnjcm.nl/en/dossiers/profiling-and-syri/ & https://pilpnjcm.nl/wp-content/uploads/2019/08/EN-Subpoena-SyRI.pdf

to The Public Interest Litigation Project, SyRi works in a which may disadvantage certain protected groups[172] –

> SyRI is only used in poor districts
>
> SyRI is currently only being used in the following cities and districts: Capelle aan den IJssel, Eindhoven, Schalkwijk in Haarlem and Hillesluis and Bloemhof in Rotterdam. These are all poor municipalities, or the poorest neighbourhoods in a municipality. In addition, there is an above-average percentage of non-Western migrants living in Schalkwijk, Hillesluis and Bloemhof. According to the PILP-NJCM, this could indicate the possible discriminatory use of SyRI with regard to people with a low income and on the grounds of ethnicity.

In early 2020, the Court of the Hague ruled that SyRI breached Article 8 of the European Court of Human Rights because of the way in which the AI system, which lacked transparency and appropriate safeguards, collated and processed such broad personal data.[173] This decision is not being appealed.[174]

Other branches of central government are also using algorithms to make decisions concerning tax and social security benefits. Here is an excerpt from the PhD thesis of Marlies van Eck on semi-automated decision making, which was provided by the Netherlands Institute for Human Rights[175] -

> In the Netherlands, the execution of legislation by the central government is divided over several specialized agencies that operate at national level. Some of them make administrative decisions that have financial and legal impact on individual citizens. For instance, the Employee Insurance Agency (UWV) makes decisions regarding applications for unemployment benefits according to the Unemployment Insurance Act (WW), the Tax and Customs Administration (Belastingdienst) decides on the annual tax returns, or the Social Insurance Bank (SVB) makes decisions on applications on child benefits. At a macro-economic level these agencies play crucial role in reallocating financial means between citizens in the Netherlands. Tax revenues are transferred from the tax administration via Treasury to agencies that can spend it on social benefits. The collaboration between employers, UWV, Belastingdienst, enables the reallocation of 156 billion euro (i.e. 60% of the State budget / treasury) and the transfer of 20 billion data per year. They call their collaboration the 'aorta' of the Dutch economy.
>
> These executive branches of public administration have 'outsourced' their tasks to computers and electronic networks over the years. Technology enabled the agencies to delegate legal administrative decision making to computers. Tasks that require

172 See https://pilpnjcm.nl/en/dossiers/profiling-and-syri/

173 See https://ekker.legal/2020/02/02/syri/

174 See https://www.openglobalrights.org/landmark-judgment-from-netherlands-on-digital-welfare-states/

175 See https://pure.uvt.nl/ws/portalfiles/portal/20399771/Van_Eck_Geautomatiseerde_ketenbesluiten.pdf, English summary on pages, 439-448.

> calculations for large numbers of citizens, such as establishing a financial relationship between administration and a citizen, are automated. Arising technological opportunities to share information made it possible for the different government agencies to interlink their systems and share citizen's personal data within supply chains and information networks. Different government agencies became able to build their legal administrative decisions based on data that is already processed by another agency. If we state that administrative decision making is the core business of public administration, then we conclude that computers execute this core business.

The Netherlands Institute for Human Rights also stated that there has been media attention in the Netherlands on the extent to which the municipalities are using algorithms in order to make decisions.[176] Additional information on these ADM processes has been collated by AlgorithmWatch with links to the supporting documentation (in Dutch). This summary is particularly helpful[177] -

> In recent times, on the lower administrative levels (especially in municipalities), a broad range of data-driven or algorithm-based initiatives have seen the light of day. It goes beyond the stretch of this Report to give a detailed overview of all developments at this point, but over recent years many municipalities have, for example, launched smart city initiatives. These initiatives collect a broad range of data from a variety of sources and for a variety of reasons, such as improving safety in entertainment districts and crowd control, but also to regulate air quality and to solve mobility issues. An important development in this regard is the creation by a coalition of (larger) municipalities in collaboration with industry and scientists of the NL Smart City Strategie in January 2017.
>
> ADM is also used in some municipalities to prevent and detect truancy and early school-leaving. This is done by using algorithms that help decide which students will be paid a visit by a school attendance officer. Similar initiatives exist to detect child abuse and/or domestic violence.
>
> Other than using System Risk Indication (see below), some municipalities have also developed their own initiatives that revolve around the use of algorithms to detect welfare fraud. These programmes take into account data such as dates of birth, family composition, paid premiums and benefits history, as well as data from the Tax and Customs Administration, Land Registry and the Netherlands Vehicle Authority. Municipalities thus hope to increase the chances of identifying people committing welfare fraud.
>
> An overview of initiatives can be found in the 2018 Report Datagedreven sturing bij gemeenten (Data driven steering in municipalities) [NL 34], which was initiated by the Association of Netherlands Municipalities. The Report urges municipalities to share knowledge, and encourages them to cooperate in the roll-out of new initiatives.

176 Equinet survey research.

177 See https://algorithmwatch.org/en/automating-society-netherlands/

There is also a concern, which is currently being researched as explained below, that algorithms are being used within Dutch recruitment processes although the precise extent of this phenomenon is currently unknown.

Predictive policing is also being used in the Netherlands. For example, a specialist system has been developed by the national police called Criminaliteits Anticipatie Systeem (Crime Anticipation System). This is how the system was described in 2017 by the Holland Times[178] -

> Today, the Dutch police use this predictive algorithm extensively. It is called Crime Anticipation System (CAS) and continuously provides updates based on the current local conditions. Consider bike thefts, for instance. The Amsterdam police was informed that the probability of their occurrence shot up after 10 pm in a particular neighbourhood. This resulted in resources being deployed accordingly. Such trials in Amsterdam were extremely encouraging, and if outskirts were included, statistics showed that over 30% of thefts were committed in the zones predicted by the algorithm. This led to the technique being tested and validated in the rest of the Netherlands.
>
> The Netherlands Scientific Council for Government Policy (WRR) operates independently and provides the government crucial advice of a range of matters. WRR has investigated predictive policing thoroughly and made important recommendations. Typically, each zone analysed is 125 by 125 metres in area and the predictions hold valid for a two-week period. Water and fields are not included in these zones. The CAS focuses on high impact crimes like home burglaries, assaults and street robberies. Simultaneously, it is being extended to others such as pickpocketing and business burglaries.

In 2018, a small pilot was also started in the Netherlands to examine the extent to which ADM could be used within the judicial system to conduct first *"sifts"*.[179] There is no further identifiable information about how this piloting process progressed.

Poland

AlgorithmWatch states that algorithms are used extensively in Poland for the allocation of judges to cases, the allocation of children to schools, the profiling of unemployed people by the government, to identify fraud within the national health service, to motivate employees and to detect financial fraud.[180]

178 See https://www.hollandtimes.nl/articles/national/predicting-crime-using-big-data/

179 See https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Oost-Brabant/Nieuws/Paginas/Rechtbank-en-universiteit-stellen-leerstoel-Data-Science-Rechtspraak-in.aspx?pk_campaign=pvs

180 See https://algorithmwatch.org/en/automating-society-poland/

Slovenia

In Slovenia, algorithms are used in border control, to grant loans and mortgages, to detect learning problems in schools, to assess insurance risk and recommend insurance products and the government uses ML to detect tax fraud.[181]

Spain

According to AlgorithmWatch, AI is being used extensively in Spain in a whole range of scenarios, from improving crop management to monitoring unemployed individuals so as to allocate job offers and training, to detect whether calls to the police are fabricated complaints, to assess the behaviour of prisoners, to identify when it is necessary to provide pre-emptive support to elderly people before an emergency arises or a request is made, to detect illegal short term letting, to predict crimes, to assist legal professionals to assess their cases, to avoid financial fraud, to diagnose bipolar disorder, to process clinical records and to diagnose diabetic retinopathy.[182] ML is also being used to assess the risk of violence in teenagers. This is how the system is described by AlgorithmWatch[183] –

> ... the Structured Assessment of Violence in Youth (SAVRY) ... system is used in forensic criminology and it was developed for assessing the risk of violence in adolescents (aged 12-18), but it was also seen to be effective in predicting the risk of general criminal recidivism. SAVRY plays a role in individual lives, and it influences the youth crime rate, as it can be used in intervention planning, such as clinical treatment plans or release and discharge decisions.

Sweden

AI is used in Sweden to automatically process compensation claims for delayed flights and travel plans, credit score assessment, to plan routes for lorries, to automatically lodge home insurance claims,to detect dyslexia, in recruitment and other personal administrative systems.[184] AI is also used to determine the eligibility of social security benefits within *"The Trelleborg model"* which is described by AlgorithmWatch as follows –

> Since 2017, Trelleborg has automated parts of its decision-making when it comes to social benefits. New applications are automatically checked and cross-checked with other related databases (e.g. the tax agency and unit for housing support). A decision

181 See https://algorithmwatch.org/en/automating-society-slovenia/

182 See https://algorithmwatch.org/en/automating-society-spain/

183 See https://algorithmwatch.org/en/automating-society-spain/

184 See https://algorithmwatch.org/en/automating-society-sweden/

is automatically issued by the system. The number of caseworkers has been reduced from 11 to 3 and the municipality argues that they have considerably reduced the number of people receiving social benefits. They have been heading a pilot project to export their automation model to 14 additional municipalities and have received several innovation prizes. However, applicants and citizens have not been explicitly informed about the automation process.

Recently, a municipality in Sweden received significant media attention after it received a fine for breaching the GDPR by monitoring the attendance of high school students via tags and FRT without ensuring adequate consent.[185]

UK

In the UK there is widespread use of algorithms, AI and ML.

One important area relates to the use of complex AI systems supported by ML within the criminal justice system. One tool which has received a reasonable amount of media attention is the Harm Assessment Risk Tool (HART) which has been utilised, since 2017, by Durham Constabulary. It deploys a ML algorithm to classify individuals according to their *"risk"* of committing violent or non-violent crimes in the future.[186] This classification is created by examining an individual's age, gender and postcode (which can be a proxy for race). The *"risk"* rating generated by HART is being used by custody officers to make significant decisions concerning people's liberty, for example, whether an individual should be permitted to access an *"out of court"* disposal programme.[187]

FRT is also being used widely by police forces in the UK in order to identify individuals on *"watch lists"*. The Law Society in the UK has recently discussed the widespread use of FRT by police forces in the UK. It stated that the following organisations were deploying FRT: London Metropolitan Police, South Wales Police, Leicestershire Police.188 Equally, FRT is being utilised by the Home

185 See https://www.datainspektionen.se/globalassets/dokument/beslut/facial-recognition-used-to-monitor-the-attendance-of-students.pdf

186 Babuta, A., Oswald, M. and Rinik, C., 2018. Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges; see https://rusi.org/sites/default/files/20180329_rusi_newsbrief_vol.38_no.2_babuta_web.pdf

187 A comprehensive review of the use of this type of "predictive policing" technology is outlined in Hannah Couchman's Report produced for Liberty entitled, "Policing by Machine – Predictive Policing and the Threat to Our Rights"; see https://www.libertyhumanrights.org.uk/wp-content/uploads/2020/02/LIB-11-Predictive-Policing-Report-WEB.pdf

188 "Algorithms in the Criminal Justice System", the Law Society, June 2019: https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/

Office in the UK to process passport applications online.189 Private sector organisations are also using FRT. Over 2019, the media focused on news that property companies in Kings Cross in Central London were using this technology190 leading the UK's Information Commissioner's Officer to launch an inquiry.191 The entertainment industry is even deploying FRT, with a bar in the UK using it to form "intelligent queues" for drinks.192

ADM is also being utilised by the UK government in order to assist with important decisions relating to immigration status specifically, the "*Settled Status scheme*"[193].

AI is also being used in the UK to predict criminal behaviour to focus resources. Recent research conducted by Sky News and Cardiff University identified that 53 local authorities are using algorithms to predict behaviour. This included Bristol City Council which uses its "*Bristol Integrated Analytics Hub*" to analyse data relating to benefits, school attendance, crime, homelessness, teenage pregnancy and mental health from 54,000 local families to predict which children could suffer from domestic violence, sexual abuse or going missing.[194] This research also identified that Kent Police now only investigates 40% of cases, as opposed to 75%, on the basis of predictive algorithms.

Local authorities in the UK are also using AI and ML in relation to Risk Based Verification (RBV) as part of determining eligibility for benefits.[195]

Moreover, it seems that the DWP is committed to expanding the scope of ADM. According to one report, as of February 2019, the department had implemented automation in 15 processes and was planning to launch another 11 automations over the course of 2019.[196] This shift towards the automation of benefits led

189 See https://www.newscientist.com/article/2219284-uk-launched-passport-photo-checker-it-knew-would-fail-with-dark-skin/

190 See https://www.theguardian.com/uk-news/2019/aug/12/regulator-looking-at-use-of-facial-recognition-at-kings-cross-site

191 See https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/08/statement-live-facial-recognition-technology-in-kings-cross/

192 See https://www.telegraph.co.uk/technology/2019/10/11/trip-worlds-first-ai-bar-proves-facial-recognition-pubs-will1/

193 See https://ai-lawhub.com/april-2019/

194 See https://news.sky.com/story/the-controversial-tech-used-to-detect-problems-before-they-happen-11649080

195 See https://ai-lawhub.com/april-2019/

196 See https://www.publictechnology.net/articles/features/ai-week-dwp-reaps-robotic-rewards

the Special Rapporteur on extreme poverty and human rights, Philip Alston, to explain in his Report on the UK dated 12 April 2019 that197-

> Benefit claims are made online and the claimant interacts with authorities primarily through an online portal. The British welfare state is gradually disappearing behind a webpage and an algorithm, with significant implications for those living in poverty.

According to the UK's Independent Chief Inspector of Borders and Immigration, a form of RBV has also been used in the UK in relation to processing visa applications by which applications are streamed into *"super green", "green", "amber"* or *"red"* category depending on perceived risk as determined by a data set.[198]

Some law enforcement authorities in the UK are also collaborating with academics and data scientists to develop tools to tackle homelessness.[199] BAE systems has recently partnered with Gloucestershire Constabulary in a £250,000 pilot project to combine data from police, social care, education and health systems in order to automatically identify cases where child protection input is required.[200]

The Law Society reported in 2019 that algorithms were being used by the Ministry of Justice as part of a *'digital reporting tool'* to manage offenders. The tool analyses live data on prison inmates' conduct during incarceration which then informs decisions such as which prison or wing an individual is assigned to. The Law Society has stated that[201] –

> The data that is used includes things like involvement in assaults, disorder and seizures of contraband such as drugs and mobile phones - as well as demographic information and location history. Details about new incidents are logged on the database shortly after they take place, which can result in new scores being generated regularly. Inmates' „scores" can change to take into account improvements or deteriorations in their behaviour.

197 See https://undocs.org/en/A/HRC/41/39/Add.1

198 See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/631520/An-inspection-of-entry-clearance-processing-operations-in-Croydon-and-Istanbul1.pdf

199 See https://www.localgov.co.uk/Councils-join-pilot-to-find-data-driven-solutions-to-homelessness/48265

200 See https://www.newscientist.com/article/2219708-data-trial-identifies-vulnerable-children-who-may-otherwise-be-missed/?utm_medium=SOC&utm_source=Twitter#Echobox=1570892317

201 See https://www.lawsociety.org.uk/support-services/research-trends/algorithms-in-the-justice-system/

Whilst the precise extent to which AI is being used in the UK in relation to personal insurance is unknown, a recent Report by the UK's "Centre for Data Ethics and Innovations" (CDEI) highlighted a number of potential ways in which it might be deployed.[202] A key passage is as follows –

> These new AI systems are expected to alter at least four dimensions of the industry:
>
> Onboarding – AI is already used to identify new customers and speed up the process of providing quotes. Insurers and price comparison websites can make use of AI-powered online advertising to segment consumers and target adverts at those more likely to be looking for a policy. Insurers have also developed chatbots that use natural language processing and generation to answer customer queries and offer quotes, including via social media platforms like Facebook Messenger. The insurer Lemonade claims its chatbot can provide a personalised policy in just 90 seconds.
>
> Pricing – AI can improve pricing by finding new patterns between personal characteristics and specific risks (e.g. between someone's credit score and the quality of their driving).3 Combined with real-time collection of data through sensors, the use of AI opens the door to hyper personalised risk scores, allowing premiums to be based on people's actual behaviour (e.g. their exercise regime), not just the risk profile of a category to which they belong (e.g. their age group, postcode or family health conditions). A related use of AI is for customer retention, with insurers modelling the minimum benefit it would take for customers to renew their policy.
>
> Claims management – AI can improve claims management by identifying fraudulent behaviour or predicting it before a claim is made. Hanzo has created AI tools that can trawl social media sites including Facebook and Twitter for corrupting evidence, such as messages that reveal someone was in a different location to the one they say they were at the time of an accident. AI can also be used to undertake damage assessments. UK-based Tractable has created an AI package that can review pictures taken at the scene of a car crash and provide an instant estimate of repair costs.4 At the back-end of insurance firms, AI can be deployed to extract relevant claims information from the bundles of written evidence passed onto insurers, including medical invoices and police reports.
>
> Advising – AI can be used to advise customers on how to avoid risks. AXA's "Xtra" health app includes a chatbot that can suggest ways for policyholders to meet fitness and nutrition goals. US tech company Cape Analytics combines machine learning software with aerial images of people's houses to analyse the quality of their rooftops - information that can then be channelled to customers to help them spot and repair damage before it worsens. In the future, insurers may be able to use AI to steer the behaviour of policyholders in real time, for example by notifying drivers of different travel routes that are known to be safer.

British companies have also been experimenting with using AI in order to assist recruitment exercises through analysing videos of job interviews to determine

202 CDEI "Snapshot Paper - AI and Personal Insurance" (September 2019); see https://www.gov.uk/government/publications/cdei-paublishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-ai-and-personal-insurance

a candidate's manner by examining mouth and eye movements and tone of voice.[203]

The UK's National Health Service has created a new unit called NHSX which will harness technology including AI to improve patient care.[204]

AI tools are being deployed into some schools in the UK to monitor the mental health of students including predicting self-harm, eating disorders and drug abuse.[205]

In the UK, algorithm-powered AI is being used within chatbots to allow individuals to report bullying and harassment.[206]

203 See https://www.telegraph.co.uk/technology/2019/10/06/does-avivas-facial-expression-technology-experiment-say-future/?WT.mc_id=tmg_share_tw and also https://www.telegraph.co.uk/news/2019/09/27/ai-facial-recognition-used-first-time-job-interviews-uk-find/?WT.mc_id=tmg_share_tw

204 See https://www.gov.uk/government/news/nhsx-new-joint-organisation-for-digital-data-and-technology and https://www.gov.uk/government/speeches/embracing-ai-and-technology-to-improve-patient-outcomes

205 See https://news.sky.com/story/artificial-intelligence-being-used-in-schools-to-detect-self-harm-and-bullying-11815865

206 See https://www.telegraph.co.uk/news/2019/09/16/barristers-get-app-report-widespread-problem-bullying-sexual/?WT.mc_id=tmg_share_tw

APPENDIX 2: ARTIFICIAL INTELLIGENCE INITIATIVES IN EUROPE

Some of the most important examples of European initiatives which seek to address potentially discriminatory AI systems are outlined here.

Pan-European level

Council of Europe

The work of the Council of Europe (CoE) in this field is very important. Its programme of work in this field is already well-developed and should be actively monitored by Equinet's Members.

In 2018, the CoE published an excellent standard-setting document written by Prof. Frederik Zuiderveen Borgesius called *"Discrimination, AI, and algorithmic decision-making"*.[207]

Since then, the CoE has developed a website dedicated to addressing human rights issues raised by AI.[208] Its aim is to move towards an application of AI based on human rights, the rule of law and democracy. It has a variety of committees examining AI, including a dedicated "Ad Hoc Committee on Artificial Intelligence" (CAHAI). The CAHAI will examine the feasibility of a legal framework for the development, design and application of AI, based on CoE's standards on human rights, democracy and the rule of law.

Marija Pejčinović Burić, Secretary General of the CoE, recently underlined the significance of its work programme in determining what more must be done to protect these rights, saying that she –

> ...look[s] forward to the outcome of the work of the Ad hoc Committee on Artificial Intelligence (CAHAI), mandated by the Committee of Ministers to "examine the feasibility and potential elements on the basis of broad multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law."

207 Zuiderveen Borgesius, F., 2018. Discrimination, artificial intelligence, and algorithmic decision-making; see https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

208 See https://www.coe.int/en/web/artificial-intelligence/home

The CoE also has a "Committee of experts on Human Rights Dimensions of automated data processing and different forms of artificial intelligence" (MSI-AUT) which will draw upon the existing CoE standards and the relevant jurisprudence of the ECtHR with a view to the preparation of a possible standard setting instrument on the basis of the study on the human rights dimensions of automated data processing techniques (in particular algorithms and possible regulatory implications).[209]

In 2019, the Office of the CoE Commissioner for Human Rights produced a practical guide called *"Unboxing Artificial Intelligence: 10 steps to protect human rights"*.[210]

On the 8th April 2020, the CoE adopted Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems. This important document adopts specific "Guidelines on addressing the human rights impacts of algorithmic systems", directed to both states and the private sector.[211]

European Union Agency for Fundamental Rights

In September 2018, the FRA published its Report "*#BigData: Discrimination in data-supported decision making*" which explained the ways in which AI and algorithms can discriminate alongside analysis of the principle of transparency and the role of the GDPR in creating accountability.[212]

In December 2018, the FRA published a new Report entitled, "*Preventing unlawful profiling today and in the future: a guide*" which examined the interplay between discrimination and data protection in the context of profiling.[213]

In June 2019, the FRA released its paper, "*Focus paper: Data quality and artificial intelligence: mitigating bias and error to protect fundamental rights*" which usefully addresses the problem of systems based on incomplete or biased data and shows how they can lead to inaccurate outcomes that infringe on people's fundamental rights, including discrimination.[214]

---

209 See https://www.coe.int/en/web/freedom-expression/msi-aut

210 See https://edoc.coe.int/en/artificial-intelligence/7967-unboxing-artificial-intelligence-10-steps-to-protect-human-rights.html

211 See https://search.coe.int/cm/pages/result_details.aspx?ObjectId=09000016809e1154

212 See https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf

213 See https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_en.pdf

214 See https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf

FRA also released in 2019 a Report entitled "*Facial recognition technology: fundamental rights considerations in the context of law enforcement*" which examines the data protection and discrimination consequences of FRT.[215]

The FRA has also collated a very detailed record of the resources currently available.[216]

European Commission (EC)

The EC set up a High-Level Expert Group on Artificial Intelligence (AI HLEG) in June 2018, as part of its AI Strategy.[217] AI HLEG produced a key document entitled "The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)"[218] on the 8 April 2019

These Ethics Guidelines are hugely influential as they explain the ethical principles and values that AI HLEG advise should underpin all development and use of AI systems, so all Equinet's Members need to be fully aware of them. They are based on the following principles –

> Human agency and oversight: AI systems should enable equitable societies by supporting human agency and fundamental rights, and not decrease, limit or misguide human autonomy.
>
> Robustness and safety: Trustworthy AI requires algorithms to be secure, reliable and robust enough to deal with errors or inconsistencies during all life cycle phases of AI systems.
>
> Privacy and data governance: Citizens should have full control over their own data, while data concerning them will not be used to harm or discriminate against them.
>
> Transparency: The traceability of AI systems should be ensured.
>
> Diversity, non-discrimination and fairness: AI systems should consider the whole range of human abilities, skills and requirements, and ensure accessibility.
>
> Societal and environmental well-being: AI systems should be used to enhance positive social change and enhance sustainability and ecological responsibility.
>
> Accountability: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes.

215 See https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper.pdf

216 See https://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights/ai-policy-initiatives

217 See https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

218 See https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines

On the 8 April 2019, the EC also explored the human rights implications of artificial intelligence in its communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, entitled "Building Trust in Human Centric Artificial Intelligence".[219]

In June 2019, the AI HLEG published its second paper entitled "Policy and investment recommendations for trustworthy Artificial Intelligence".[220] This paper repeatedly emphasised the importance of building a FRAND (fair reasonable and non-discriminatory) approach, and proposed regulatory changes, arguing that the EU –

> Adopt a risk-based governance approach to AI and an ensure an appropriate regulatory framework Ensuring Trustworthy AI requires an appropriate governance and regulatory framework. We advocate a risk-based approach that is focused on proportionate yet effective action to safeguard AI that is lawful, ethical and robust, and fully aligned with fundamental rights. A comprehensive mapping of relevant EU laws should be undertaken so as to assess the extent to which these laws are still fit for purpose in an AI-driven world. In addition, new legal measures and governance mechanisms may need to be put in place to ensure adequate protection from adverse impacts as well as enabling proper enforcement and oversight, without stifling beneficial innovation.

In the summer of 2019, the EC said that it would launch a pilot phase involving a wide range of stakeholders.[221] Following the pilot phase, in early 2020, the AI expert group will review the assessment lists for the key requirements, building on the feedback received. Building on this review, the EC proposes to evaluate the outcome and propose any next steps.

At the beginning of Chapter 1, the commitment made by the new EC President was noted. The authors expect that the new Commissioner Margrethe Vestager, Executive Vice-President for a Europe fit for the Digital Age will take this forward early in 2020. She explained her intended approach in answers given to a Questionnaire from the European Parliament on the 8 October 2019, thus[222] -

219 See https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence

220 See https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence

221 See https://ec.europa.eu/cyprus/news/20190408_2_en

222 See https://ec.europa.eu/commission/commissioners/sites/comm-cwt2019/files/commissioner_ep_hearings/answers-ep-questionnaire-vestager.pdf.

> Artificial intelligence can serve us in many sectors of the economy, such as health, transport, communication and education. It can enable a wide-scale automation of decisions and processes that has an enormous potential to increase quality, efficiency and productivity. It will impact many aspects of our lives, from self-driving cars to improved medical procedures. At the same time, this technology, which is based on self-learning and self-improving algorithms, can raise many policy issues, for instance issues such as accountability or social acceptance.
>
> In this context, the President-elect entrusted me with the responsibility to coordinate work on a European approach on Artificial Intelligence, including its human and ethical implications. This effort will feed into the broader work stream on industrial policy and technological sovereignty, as we must ensure that European citizens and companies can reap the benefits of this technology as well as shape its development.
>
> Our work will also build on the existing policy achievements, in particular the ethical guidelines that were adopted in June 2019. Their application is currently being tested. It is therefore our intention in the first 100 days of the new Commission to put forward proposals developing the European approach for Artificial Intelligence.
>
> Our objective is to promote the use of Artificial Intelligence applications. We must ensure that its deployment in products and services is undertaken in full respect of fundamental rights, and functions in a trustworthy manner (lawful, ethical and robust) across the Single Market. This approach must provide regulatory clarity, inspire confidence and trust, and incentivise investment in European industry. It should improve the development and uptake of Artificial Intelligence in the EU while protecting Europe's innovation capacity. As part of our approach to an overall framework for Artificial Intelligence we will also review the existing safety and liability legislation applicable to products and services.
>
> This will ensure in particular that consumers benefit from the same levels of protection independently of whether they are using traditional products or smart, digitally enabled products (e.g. smart fridge, smart watches, voice-controlled virtual assistants).
>
> Given the complexity of the issues at stake, a wide and thorough consultation of all stakeholders, including those who have participated in the pilot on implementing the ethics guidelines developed by the high-level expert group, would be required. We will look carefully at its impact across the board and make sure that our new rules are targeted, proportionate and easy to comply with, without creating any unnecessary red tape.

In February 2020, the EC published its long awaited White Paper – "On Artificial Intelligence – A European approach to excellence and trust".[223] The purpose of the White Paper is to start the process of scoping policy options which are intended to "enable a trustworthy and secure development of AI in Europe" and avoid localised regulation which would lead to "a real risk of fragmentation in the internal market, which would undermine the objectives of trust, legal certainty and market uptake".

223 See https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

According to the White Paper, the specific areas where the existing EU legislative framework could be improved are as follows:

i. Ensuring greater levels of transparency
ii. Extending EU product safety legislation to AI systems
iii. Ensuring that AI systems which change as they are utilised can be effectively policed
iv. Clarifying legal responsibility for AI systems within the supply chain
v. Extending the meaning of "safety" to capture the potential harms created by AI
vi. Introducing a risk-based approach to regulation so that intervention is proportionate
vii. Regulating the use of data sets which "train" AI
viii. Prescribing the keeping of records concerning the data set used, its accuracy and how it is used
ix. Ensuring that citizens are always informed about what AI systems can do and how they are used
x. Ensuring that citizens are informed when they are interacting with a non-human
xi. Ensuring that AI systems are accurate
xii. Guaranteeing human oversight
xiii. Creating special rules for biometric data

Alongside the White Paper, the EC also released its "Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics" which provides a more practical perspective on legislative reform.[224] The Report primarily focuses on how the General Product Safety Directive[225] and harmonised product legislation can be amended to include the regulation of AI. Much of this analysis is therefore premised on AI systems being analogous to other products such as medical devices. Whilst there are certainly some useful parallels to be drawn, there are plainly limitations to conceptualising AI as simply another type of regulated "product".

Despite this limitation, there are six important proposals in the Report which could greatly assist the regulation of AI:

224 See https://ec.europa.eu/info/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en

225 Directive 2001/95/EC of the European Parliament and of the Council of 3 December 2001 on general product safety; see https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX-:32001L0095&from=EN

xiv. Imposing an obligation on developers of algorithms to disclose design parameters and metadata of datasets.

xv. Confirmation of the principle that whoever places an AI system in the market is responsible for its safety regardless of the complexity of the supply chain.

xvi. A requirement for actors within the supply chain to co-operate with one another to ensure the safety of AI systems.

xvii. Reversal of the burden of proof in relation to harms caused by AI systems.

xviii. Requiring producers of AI systems to ensure that they are safe throughout their lifecycle rather than simply at the point of sale.

xix. The introduction of strict liability for certain products.

European Council

As early as 2018, the European Council had emphasised the importance of AI alongside the importance of analysing it within a human rights framework. Its 2019 document *"Coordinated Plan on the Development and Use of Artificial Intelligence Made in Europe"* repeats this message.[226]

European Data Protection Board

The aim of the EDPB is to contribute to the consistent application of data protection rules throughout the EU and promote cooperation between the EU's data protection authorities. The Board is established by the GDPR. As part of its work programme, it has addressed matters which relate to AI and its potential to discriminate. It has published a Report on ADM called *"Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation (2018)"*[227] and algorithms in the financial sphere entitled "New Rules for Credit Reporting Systems in the Digital Economy (2019)".[228]

National level

Legislation

No examples of countries within Equinet's jurisdiction which had enacted AI specific legislation to tackle discriminatory systems expressly could be iden-

226 See https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/en/pdf

227 See https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

228 See https://edpb.europa.eu/news/national-news/2019/new-rules-credit-reporting-systems-digital-economy_en

tified.[229] Although some legislatures are considering the issue such as France[230] and Germany[231] or are taking steps towards creating greater accountability such as Denmark, the Netherlands[232] [233] and the UK.[234] Malta has very recently introduced a certification scheme for Artificial Intelligence. [235]

Some countries in Europe have supplemented the GDPR by requiring a certain level of transparency in relation to algorithms particularly in Denmark[236], Italy,[237] and France.[238]

Artificial intelligence strategies

At a national level, many countries in Europe are actively engaging with AI from a commercial perspective. That is, formulating so called *"AI strategies"* in order to embed a technology which is perceived as bringing *significant* economic benefits.

National ethical frameworks or bodies

Many of countries which had developed commercial AI strategies, have made reference to the importance of AI being *"ethical"* (although not always non-discri-

229 There are some countries in which additional legislation has been enacted in light of the GDPR and for other data protection reasons, which do regulate automated decision making, but the authors have not been able to identify legislation which is intended to target expressly discriminatory artificial intelligence.

230 See https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

231 See https://www.bundestag.de/en/committees/bodies/study/artificial_intelligence

232 There is a motion currently being considered to create a register of algorithms used in the public sector; see https://algorithmwatch.org/en/story/kees-verhoeven-algorithm-registry/

233 Netherlands is considering whether local authorities which use algorithms will be required to report to an independent body so as to improve transparency and accountability: see https://www.nu.nl/politiek/5997764/tweede-kamer-wil-een-algoritmemeldplicht-voor-de-overheid.amp?__twitter_impression=true https://algorithmwatch.org/en/story/kees-verhoeven-algorithm-registry/

234 The government published draft "Guidelines for AI procurement" which are intended to "... help inform and empower buyers in the public sector, helping them to evaluate suppliers, then confidently and responsibly procure AI technologies for the benefit of citizens": see https://www.gov.uk/government/publications/draft-guidelines-for-ai-procurement

235 It is not clear though as to whether this scheme will examine equality issues: see https://www.maltachamber.org.mt/en/malta-first-country-in-the-world-to-launch-ai-certification-programme

236 See https://algorithmwatch.org/en/automating-society-denmark/

237 See https://algorithmwatch.org/en/automating-society-italy/

238 See https://algorithmwatch.org/en/automating-society-france/

minatory), for example, Malta[239], Lithuania[240], Portugal[241], France[242], Belgium[243], the Netherlands[244] and the Czech Republic[245]. Some countries have also developed or are intending to create AI specific ethical frameworks and /or, have formed boards to examine ethical AI and/or ethical data, such as Denmark[246], the UK[247], Italy[248], Finland[249] and the Netherlands[250]. As the OECD agreed Principles on Artificial Intelligence on the 22nd May 2019,[251] which are very similar to those of the AI HLEG, the authors expect that more countries may continue down the same path.

Germany's Federal Government set up the Data Ethics Commission on 18 July 2018. [252] It asked the Commission key questions concerning algorithm-based decision-making, AI and data. In October 2019, the Commission published its opinion (available in English).[253]

National data protection authorities

There are data protection bodies within Europe that are addressing the extent to which ADM is lawful from a data protection perspective. For example, the Data Protection Ombudsman in Finland has provided guidance to the public on their

239 See https://malta.ai/

240 See http://kurklt.lt/wp-content/uploads/2018/09/StrategyIndesignpdf.pdf

241 See https://www.incode2030.gov.pt/en/featured/minister-science-presents-strategy-artificial-intelligence-berlin & https://www.portugal.gov.pt/download-ficheiros/ficheiro.aspx?v=236848b1-fcb6-4c65-9773-292d1c5b9ad1

242 See https://www.aiforhumanity.fr/en

243 See https://www.ai4belgium.be/

244 Equinet survey research.

245 See https://www.mpo.cz/assets/en/guidepost/for-the-media/press-releases/2019/5/NAIS_eng_web.pdf

246 See https://algorithmwatch.org/en/automating-society-denmark/

247 See https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation and https://www.gov.uk/government/organisations/office-for-artificial-intelligence and https://www.gov.uk/government/publications/draft-guidelines-for-ai-procurement/draft-guidelines-for-ai-procurement

248 See https://futureoflife.org/ai-policy-italy/

249 Equinet survey research.

250 See https://www.rijksoverheid.nl/documenten/kamerstukken/2019/10/08/kamerbrief-over-ai-publieke-waarden-en-mensenrechten and https://www.tweedekamer.nl/kamerstukken/amendementen/detail?id=2019Z19084&did=2019D39751

251 See https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

252 See https://datenethikkommission.de/

253 See https://datenethikkommission.de/wp-content/uploads/191023_DEK_Kurzfassung_en_bf.pdf

rights,[254] the UK's Information Commissioner's Officer has looked at the interplay between data protection principles and discrimination[255] and the French Data Protection Commission (CNIL) has examined the ethical implications of algorithms.[256]

Auditing through impact assessments

There is also growing idea, at least within the UK[257], that organisations which deploy AI and algorithms should be compelled to conduct and publish Algorithmic Impact Assessments or audits, similar to Data Protection Impact Assessments, which demonstrate that the potential for the technology to discriminate has been assessed and minimised. This is a concept which has been publicised by international bodies such as the AI Now Institute[258], commentators[259] and adopted by Canada.[260]

Litigation

There has been some very limited litigation in Europe concerning the use of algorithms and their potential to discriminate.

In the UK, there has been a recent judicial review of the use of FRT by South Wales police force.[261] In *R v The Chief Constable of South Wales Police ex parte Bridges*[262], a charity dedicated to preventing excessive government control called Liberty brought an action against the police in relation to its use of FRT in certain public places. The case was primarily argued on privacy grounds but there was an argument concerning the Public Sector Equality Duty which exists in UK and which requires public authorities to have regard to *"the need to eliminate discrimination harassment, victimisation ..."*.[263]

---

254 See https://tietosuoja.fi/en/have-you-been-subjected-to-a-decision-based-solely-on-automated-processing

255 See https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

256 See https://www.cnil.fr/en/algorithms-and-artificial-intelligence-cnils-report-ethical-issues

257 See https://zenodo.org/record/3237865#.XrQSyT90zD7

258 See https://ainowinstitute.org/aiareport2018.pdf

259 See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456224

260 See https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/algorithmic-impact-assessment.html

261 The authors understand that this matter is now proceeding to the Court of Appeal.

262 [2019] EWHC 2341. The full judgment can be accessed at https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf

263 Section 149 of the Equality Act 2010.

In 2017, South Wales Police undertook an initial assessment as to whether deploying FRT could lead to direct discrimination but apparently it omitted to examine whether indirect discrimination could occur.[264] Mr Bridges argued that in those circumstances there was a breach of the Public Sector Equality Duty. The court was critical of this submission, noting that it had *"an air of unreality"* because *"there is no firm evidence that the software does produce results that suggest indirect discrimination"*.

However, what the court did not grapple with sufficiently was that there was a lack of *"firm evidence"* due to a lack of transparency within the system deployed by the police. Regrettably therefore, this case did not usefully examine the equality implications of FRT.

In the Netherlands, various interest groups and individuals are litigating the use of the SyRI system as noted above.[265] It was argued that SyRI was in breach of the right to private life, the right to privacy, the GDPR and the right to an effective remedy due to the lack of transparency around the algorithm deployed by the state. As part of the arguments concerning transparency, it was argued that the system could not be interrogated so as to ensure that discrimination is not occurring which is contrary to public law principles. Here is an excerpt from the case against the government –

> 5.5 The arrangement for SyRI does not meet the foreseeability requirement in several respects. ...
>
> Risk models are secret
>
> 5.28 In the fourth place, it remains unforeseeable for the citizen at all times how SyRI will be deployed in a specific project, because the risk model remains secret. Complainants have asked in their Wob-request to be provided with the risk model. This part of the Wob-request was rejected ...
>
> 5.29 So the risk model which is used to analyse the collection of data is not disclosed at all. This is objectionable in the first place because the risk model cannot be assessed in this manner, for example against the ban on discrimination (also see marginal nos. 5.44). Discrimination based on risk models is an obvious risk. A possible distinction between various groups of citizens must be objectively justified and open to assessment against verifiable information. In the second place, citizens cannot gauge in any manner when a risk notification may be made.

264 See the judgment, op. cit., at [151] – [152].

265 A copy of the subpoena which sets out the arguments which are being advanced is available here: https://pilpnjcm.nl/wp-content/uploads/2019/08/EN-Subpoena-SyRI.pdf

When the judgment was handed down on 5 February 2020,[266] the Dutch court expressly recognised that the Netherlands government has a legitimate interest in ensuring that benefits are paid to the correct people and that fraud is detected. It also stated that the Government should use technology in order to more accurately detect fraud. However, the court also said that the right to privacy needed to be carefully protected as new technologies, which exploit big data, are deployed, –

> However, the development of new technologies also means that the right to the protection of personal data is increasingly important. The existence of adequate legal privacy protection in the exchange of personal data by (government) bodies contributes to the trust of the citizen in the government, just as the prevention and combating of fraud does. As NJCM et al. Rightly states, it is plausible that in the absence of sufficient and transparent protection of the right to respect for private life a ' chilling effect ' will occur. Without confidence in adequate privacy protection, citizens will want to provide information less quickly or there will be less support for it.

The court proceeded to find that Article 8 of the European Convention of Human Rights was breached by SyRI, as summarised here –

> The court compared the content of the SyRI legislation in the light of the purposes that this legislation serves against the breach of private life that the SyRI legislation makes. It is of the opinion that the legislation does not comply with the 'fair balance' that must exist under the ECHR between the social interest that the legislation serves and the violation of the private life that the legislation produces in order to be able to speak. about a sufficiently justified breach of private life. In doing so, the court takes into account the fundamental principles on which data protection under Union law (the Charter and AVG) is based, in particular the principles of transparency, the purpose limitation principle and the principle of data minimization. She believes that the legislation regarding the use of SyRI is insufficiently clear and verifiable. It is for that reason that the court will declare Article 65 of the SUWI Act and Chapter 5a of the SUWI Decree to be incompatible with this judgment on grounds of conflict with Article 8, paragraph 2 of the ECHR.

The specific features of SyRI which led the Court to conclude that Article 8 had been breached were, broadly speaking, as follows –

a. The sheer breadth and scope of the data processed: [6.50].
b. The use of machine learning to analyse and make links within data: [6.50].
c. People do not necessarily know whether their data is being processed and if so, the outcome of any analysis: [6.54].
d. It created "risk reports" on individuals which could have significant personal consequences: [6.60].

266 See https://ekker.legal/2020/02/02/syri/

e. There were insufficient safeguarding mechanisms within SyRI to protect individuals: see e.g. [6.72].

f. In particular, the opacity within the system made verifying its processes near impossible: see e.g. [6.90].

Importantly, the Government sought to "downplay" the sophistication of the SyRI system seeking to portray its algorithmic capability as relatively basic and asserted that it did not utilise machine learning at all ([6.48] – [6.49]). However, it also declined to provide information to verify these claims on the basis that disclosure would allow citizens to cheat the system. In those circumstances, and on the information, which was available to it, the Court broadly preferred the claimants' presentation of the SyRI system leading it to conclude that Article 8 had been breached.

The notion that SyRI discriminates against citizens was also assessed by the Court. It acknowledged that SyRI had the potential to discriminate finding that –

> ... given the large amounts of data that are eligible for processing in SyRI, including special personal data, and the fact that risk profiles are used, there is a risk that the use of SyRI will inadvertently make connections based on bias, such as a lower socio-economic status or an immigration background ...

Whilst the Court did not find that discrimination was occurring, it did conclude that the possibility of discrimination combined with an absence of transparency fortified its conclusion that Article 8 had been breached: [6.95].

In Finland, the National Non-Discrimination and Equality Tribunal was asked by the Non-Discrimination Ombudsman to adjudicate upon the credit scoring process described in Appendix 1 above. The Tribunal concluded that discrimination had occurred because credit had been refused because of the individual's place of residence, gender, age and language.[267] As a result of this decision, the Non-Discrimination Ombudsman recommended to the Financial Supervisory Authority that it should evaluate the credit scoring process of financial institutions from a non-discriminatory perspective and proposed enhanced co-operation between the two bodies.[268]

267 See https://www.yvtltk.fi/en/index/opinionsanddecisions/decisions & https://www.syrjinta.fi/web/en/-/assessing-credit-rating-on-the-basis-of-statistical-data-alone-is-discrimination-credit-institutions-must-revise-their-practices

268 Equinet survey.

Academia and other expert groups

There is a growing debate within universities about the discriminatory impact of AI. By way of example, in the UK, the Alan Turing Institute[269] has been at the forefront of academic research into the implications of AI. Sandra Wachter, an academic based at Oxford University, has written extensively on the interplay between discrimination law and AI.[270]

In the Netherlands, there is academic research being undertaken into the use of discriminatory algorithms in personnel management e.g. which candidates are invited to interviews and then ultimately selected.[271] The Netherlands Institute for Human Rights will be considering the academic research, which is due in late 2019, and may start a suitable initiative to combat any discrimination.[272] Dr Marlies van Eck wrote her PhD thesis on semi-automated decision making which examines the potential for discrimination (semi)automated administrative chain decisions and legal protection.[273]

Unia in Belgium is also working with the University of Antwerp to develop a tool to detect online hate speech.[274] It is also working with the Flemish Inter-University expert group Kenniscentrum Data en Maatschappij[275] which gives it access to projects which it co-develops or advises on. Kenniscentrum Data en Maatschappij is looking at such issues as data gathering on social distancing in the campaign to address Covid – 19.[276]

Campaigning Groups

Alongside academics, numerous organisations have developed within individual countries with the aim of placing pressure on government to increase transparency and accountability around the use of artificial intelligence. In the UK, one such organisation which has been successful at raising the profile of AI related matters is Big Brother Watch.[277]

---

269 https://www.turing.ac.uk/research/research-programmes/artificial-intelligence-ai

270 See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3388639

271 Equinet survey research.

272 Equinet survey research.

273 See https://pure.uvt.nl/ws/portalfiles/portal/20399771/Van_Eck_Geautomatiseerde_ketenbesluiten.pdf, the English summary is at pages 439-448.

274 Equinet survey research.

275 See https://data-en-maatschappij.ai/

276 See https://data-en-maatschappij.ai/nieuws/survey-onze-respondenten-zijn-sterk-verdeeld-over-gebruik-van-technologie-en-persoonlijke-gegevens-in-strijd-tegen-corona

277 See https://bigbrotherwatch.org.uk/all-media/new-statesman-how-citizen-scoring-algorithms-are-being-used-in-the-uk/

APPENDIX 3: SURVEY RESULTS

Figure 1

Q4 In your country, are organisations or people currently talking about the potential for artificial intelligence and algorithmic decision making to discriminate against individuals or groups of people?

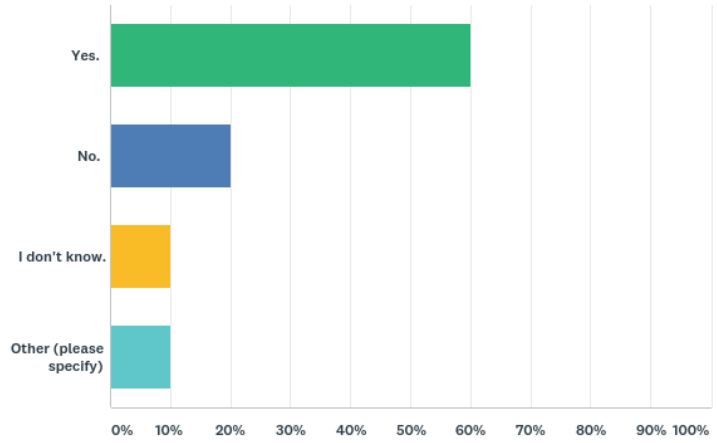


Figure 2

Q5 If yes, please explain which organisations or people are talking about the potential for artificial intelligence and algorithmic decision making to discriminate against individuals or groups of people?

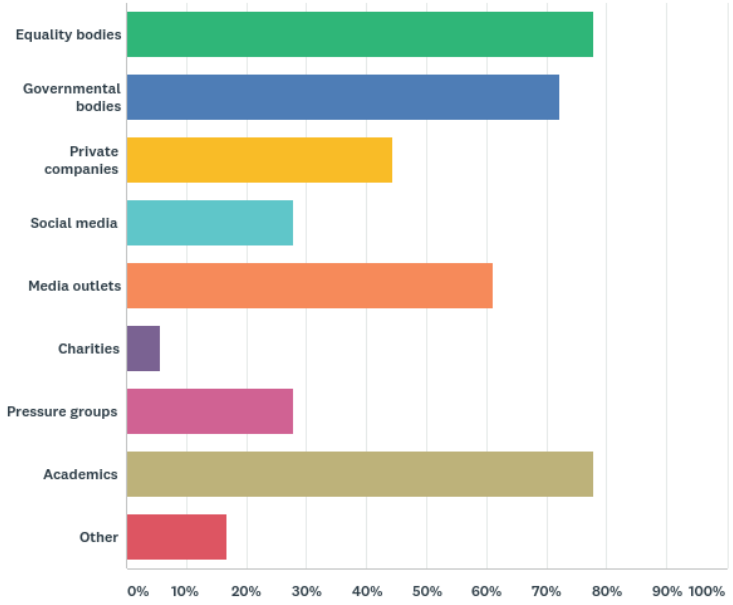


Figure 3

Q7 If there is a public debate within your country, are there certain protected characteristics where there are special concerns about discrimination?

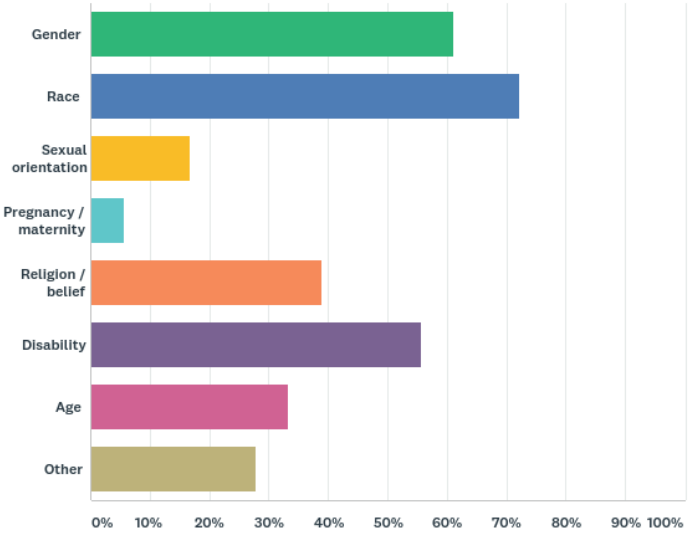


Figure 4

Q10 Is there a debate in your legislative body about the regulation of artificial intelligence and / or algorithmic decision?

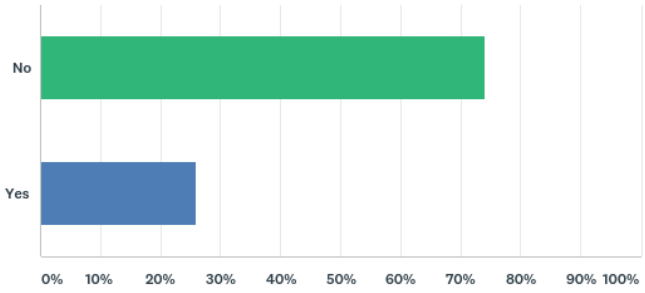


Figure 5

Q14 Has your organisation started to consider the impact of artificial intelligence and / or algorithmic decision-making on society from an equality and non-discrimination perspective?

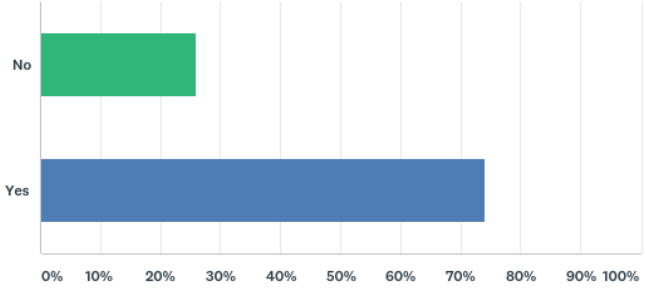


Figure 6

Q16 Has your organisation undertaken any work, or will your organisation be undertaking any work, in order to combat discrimination created by artificial intelligence or algorithmic decision making?

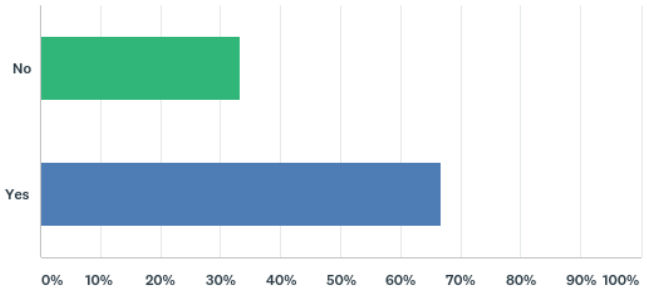


Figure 7

Q17 Is your organisation currently defining "best practice" (or will your organisation be defining "best practice") around artificial intelligence and / or algorithmic decision making from an equality and non-discrimination perspective?

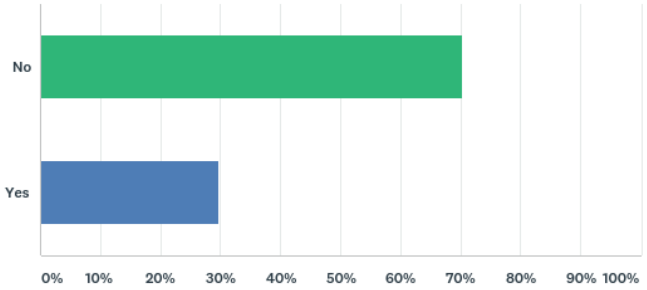


Figure 8

Q23 Is your organisation working with or thinking about working with any other organisation to examine the impact of artificial intelligence and / or algorithmic decision making on society from an equality and non-discrimination perspective?

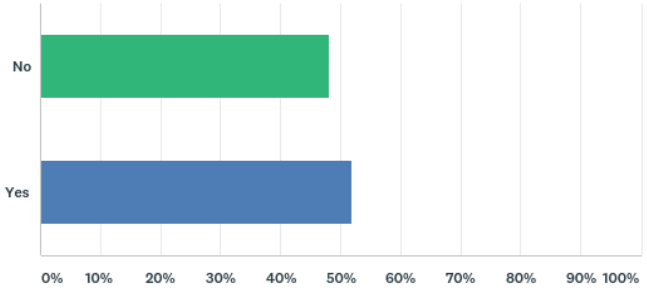


Figure 9

Q24 Is your organisation following the work of the European Union concerning artificial intelligence and how it should be regulated?

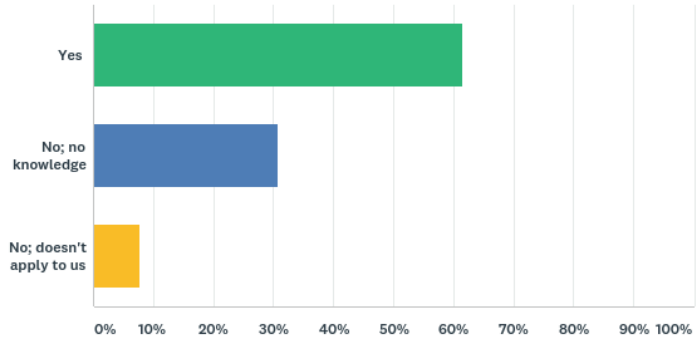
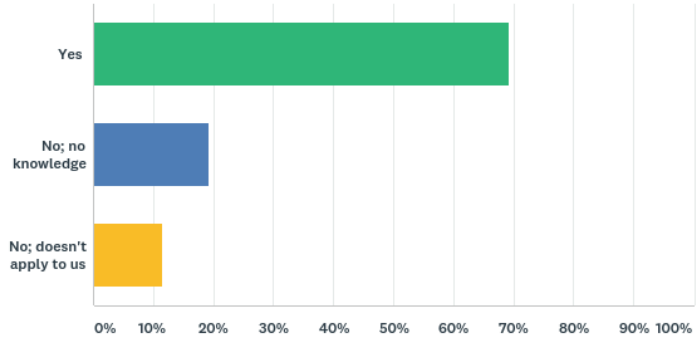


Figure 10

Q25 Is your organisation following the work of the Council of Europe concerning artificial intelligence and how it should be regulated?



ABOUT THE AUTHORS

Robin Allen QC and Dee Masters are both practising barristers (the equivalent of avocats, Anwälte, etc.) specialising in equality and human rights law. For many years they have advised a very wide range of clients in relation to ways to avoid discrimination, or to bring or defend claims, concerned with equality law. They work internationally and have conducted litigation at every level, and on every ground protected under EU law, including cases in the Court of Justice of the European Union and the European Court of Human Rights. They have also lectured and trained jurists across Europe on issues relating to equality.

They set up the AI Law Consultancy (www.ai-lawhub.com) in 2018 because they were increasingly aware of the potential dangers in the rapid proliferation of Artificial Intelligence (AI) systems and associated new technologies and the need for an ethical and clear regulatory framework to enable the best use of such systems. The website contains information about the developing regulation of AI systems and is updated frequently.

They also tweet about these developments at @AILawHub.

Jonathan Cook and Ruaraidh Fitzpatrick, colleagues in Cloisters Barristers' Chambers (www.cloisters.com), also provided helpful research at an early stage of this project.

ABOUT EQUINET

Equinet is the European Network of Equality Bodies, a membership organisation bringing together equality bodies from across Europe. Equinet promotes equality in Europe by supporting and enabling the work of national equality bodies. It supports equality bodies to be independent and effective as valuable catalysts for more equal societies.

Equality bodies are champions for the core EU value of equality and defenders of the right to non-discrimination. They are public organisations assisting victims of discrimination, monitoring and reporting on discrimination issues, and contributing to an awareness of rights and a societal valuing of equality. They are legally required to do so in relation to one, some, or all of the grounds of discrimination covered by European Union (EU) law – gender, race and ethnicity, age, sexual orientation, religion or belief, and disability.

Equinet carries out legal and policy research to inform European policy and legislative developments, and provides relevant knowledge to those interested in equality and non-discrimination in Europe, including working with relevant partners such as Cloisters.


EQUINET
European Network
of Equality Bodies