On the 31January 2020
at
Congress House
Great Russell Street
London WC1B 3LS

On the 3 February 2020
at
Dynamic Earth,
112 Holyrood Rd
Edinburgh, EH8 8AS

Michael Rubenstein Conferences Ltd

Discrimination Law in 2020

--

ARTIFICIAL INTELLIGENCE, MACHINE LEARNING,

ALGORITHMS AND DISCRIMINATION LAW:

--

THE NEW FRONTIER

--

ROBIN ALLEN QC

--

Cloisters' Robin Allen QC and Dee Masters[1] set up the AI Law Consultancy (www.ai-lawhub.com) in 2018 because we had become increasingly aware of the potential and dangers in the rapid proliferation of AI systems and associated new technologies.

Our website contains information about the developing regulation of AI systems and is updated frequently. We also tweet about these developments at @AILawHub This paper has been co-researched by Dee Masters.

The Consultancy works with business, NGOs, individuals, governments and regulators, at the interface of equality law and AI. We welcome discussion and instructions to advise on these issues.

For many years we 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. We 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.

We have lectured and trained jurists across Europe on issues relating to equality. We can be contacted through our Chambers at –

CLOISTERS

1 Pump Court,
Temple,
London EC4Y 7AA,
United Kingdom

+ 44 (0) 20 7827 4000

ra@cloisters.com or deemasters@cloisters.com

We are regulated by the Bar Standards Board.

All sites visited in January 2020.

[1] www.cloisters.com

Abbreviations

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

ADM	Automated decision making
AI	"Artificial Intelligence"; the abbreviation "AI" IS USED both specifically and generically. It will be clear from the context which meaning we intend. When we use it generically, as in for instance "AI systems", we include ML and ADM and other forms of computer algorithm derived outputs.
AI HLEG	EU's High-Level Expert Group on Artificial Intelligence
CAHAI	CoE Ad Hoc Committee on Artificial Intelligence
CDEI	Centre for Data Ethics and Innovation
CoE	Council of Europe
DWP	Department for Work and Pensions
EAD1e	The IEEE's First Edition of its 'Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems'
EC	European Commission
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
EDPB	European Data Protection Board
EHRC	Equality and Human Rights Commission
EU	European Union
FCA	Financial Conduct Authority
FRA	Fundamental Rights Agency of the European Union
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
CFREU	Charter of Fundamental Rights of the European Union
CJEU	Court of Justice of the European Union
EA	Equality Act 2010
EU	European Union
HRA	Human Rights Act 1998
IAPP	International Association of Privacy Professionals
ICO	Information Commissioner's Office
IEEE	Institute of Electrical and Electronics Engineers
IS	Intelligent Systems
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
ML	Machine learning
MSI-AUT	CoE Committee of experts on Human Rights Dimensions of automated data processing and different forms of artificial intelligence
NGO	Non-governmental organisation
RBV	Risk-Based Verification

CONTENTS

Something new for regular Discrimination Law Update Conference Attendees!

1. It is always a pleasure to speak at this annual conference which reviews the developments of the last year and starts the consideration of what the next may bring.

2. In my chambers **Cloisters** we are constantly dealing with the most important issues in discrimination law. Please visit our website www.cloisters.com for the latest equality blogs, with contributions from our barristers who are routinely involved in the most complex and high-profile equality and discrimination cases. You can also follow us on Twitter @Cloisters or myself @RobinAllenQC for regular updates and news about discrimination law.

3. Usually I am asked to speak on the case law developing around one or more of the protected characteristics. This year is different. My topic is the very fast developing area of **Artificial Intelligence (AI)** and its impact on equality and non-discrimination. This is likely to be something new for you. You will not yet find any mention of AI on the Equality and Human Rights Commission's website nor on that of the Equality Commission for Northern Ireland.

4. In fact, a quiet revolution has been going on and it has already caused a lot of concern, so that even before Brexit is done and dusted, we are likely to see European legislation specifically directed to AI.

A quiet revolution that affects us all

5. Actually, the revolution is not quite so quiet, as the new President of the European Commission (EC), Ursula von der Leyen, understands. Thus, in the Summer of 2019, even before she was elected to the post, she wrote[2] –

> 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

6. Margrethe Vestager, the EC's Executive Vice-President and new "Commissioner for a Europe fit for the Digital Age", has been given the task for taking forward the new President's promise as she explained in answers

[2] 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

given on the 8th October 2019 to a Questionnaire from the European Parliament[3] -

> 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

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

> 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.

7. Subsequently a draft of a consultation paper from the Commission marked 12/12 has recently been leaked and is available here.[4] This suggest that there will be some important activity soon, probably starting with a consultation. I shall comment further on a part of its contents below, but I should say here that this is only a leaked draft. It does not necessarily convey current thinking, and my information is that it is likely to be inaccurate in some respects.

So, what are we talking about?

8. The short answer is AI systems; these are systems that typically concern algorithms, machine learning and automated decision-making.

Algorithms

9. *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, we have, what is commonly referred to as an *Artificial Intelligence* (AI) system

10. There is no single definition of an AI system. In 2019 EU's High-Level Expert Group on Artificial Intelligence (AI HLEG)[5] adopted this definition which introduces the concept of *machine learning* (ML) –

> 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.

[4] https://www.euractiv.com/wp-content/uploads/sites/2/2020/01/AI-white-paper-EURACTIV.pdf

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

Machine Learning

11. AI HLEG continued –

> 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).

12. AI systems are thus engaged with data, and often with huge amounts of data. This is what leads to the possibility of intelligent behaviour. This so-called intelligent behaviour and reasoning is created through *machine learning* (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. The International Association of Privacy Professionals (IAPP) has described ML as follows[6] –

> 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.

Automated decision-making

13. One common application of algorithms and AI is through *automated decision-making* (ADM) where conclusions are reached, without any direct, or with only limited, human involvement. There is also a grey area in which humans do have the last say on a decision but in practice simply apply the conclusions that ADM has produced.

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

How common are AI systems? What are they being used for?

14. It is impossible to state precisely the extent to which the use of AI systems is expanding, though there are some resources that provide fleeting insights. For instance, Stanford University has produced an online "*AI Index*" that tracks the growth of AI by numerous metrics including country, sector and even business function.[7] This Index demonstrates an explosion in AI in the past 2 to 3 years.

15. We know that the economic and business drive to accelerate the use of these technologies is immense. In February 2019 the McKinsey Global Institute calculated the potential gain for Europe in adopting AI systems as follows[8] -

> 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.

16. Moreover there is not the slightest doubt that in this age of austerity, the public sector will be using AI systems to a greater and greater extent.[9] 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 the whole of Europe and the world beyond. In September 2019, KPMG released its report, "*KPMG 2019 Enterprise AI Adoption Study into AI*" which found that 30% of the 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.[10]

Tell me a bit more as to how AI works for instance in recruitment?

17. Companies are deploying technology in a range of ways in the recruitment field. We are aware that complex AI is being used to assess candidates for roles including through automated video analysis and assessment of social media presence. In particular there is a fast-developing commercial interest in the use

[7] 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

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

[9] 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/

[10] See https://advisory.kpmg.us/articles/2019/ai-transforming-enterprise.html

of AI systems to assist with recruitment in the US, and some of these will undoubtedly be used here in Europe.

18. We know that AI is being used for all of the following tasks in relation to recruitment –

    - Robot interviews
    - Conversation analytics[11]
    - AI-powered background checks
    - Reference checking
    - Internal promotion and sideways moves
    - Assessing team strengths and bridging talent gaps
    - AI-powered talent marketplaces, and
    - Determining reward strategies

Where else should we be aware of AI systems operating in ways which could be in conflict with equality and human rights law?

19. As well as in recruitment, research by Dee Masters and myself through the AI Law Consultancy, has shown that AI systems are being used now in at least the following contexts across Europe–

    - Reducing unemployment: Algorithms are being used by governments to calibrate the level of assistance that jobseekers should receive in relation to obtaining employment and 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): This form of technology, which is dependent on ML algorithms, is being deployed by the state in many countries and in the commercial world. Algorithmwatch currently estimates that at least 10 police forces across Europe are using FRT.[12]

    - Education: Algorithms are being used in a variety of ways in the field of education, from the allocation of teachers to the monitoring of students.

    - Predictive policing: We are aware that sophisticated AI is being used to make predictions about where crimes will be committed and by whom.

[11] For instance: Software audio records the interviews you conduct, both via (video) call or in-person, and then automatically transcribes them. The interview insights you get out of this can help you understand how effective your interviewers are and how to improve where necessary: see https://harver.com/blog/ai-in-recruitment-2020/

[12] See https://algorithmwatch.org/en/story/face-recognition-police-europe/

- Immigration / border control: We are aware that AI is being used to make decisions about immigration status and to control and borders.
- Financial products: Algorithms are being used in relation to credit scoring and the availability of insurance.
- Health: Technology is being used to detect and predict illnesses in a range of ways that seems to be growing daily. This is likely to be used to assist in developing preventative medicine measures which, of course, must be equally and fairly distributed, regardless of a person's protected characteristics, of race, gender, disability and so on.
- Social advantages: Algorithms are being utilised to make decisions concerning eligibility for social welfare. Governments have adopted these processes in order both to reduce the cost of determination and to try to increase predictability.
- Child welfare: Some countries are deploying AI to assess the risk of children requiring state interventions in order to protect their welfare.
- Justice and criminal justice system: Many countries are using or contemplating using AI systems in the justice and criminal system.
- Fraud detection: AI is being used to predict which individuals might be defrauding the state in several countries and the AI Law Consultancy expects that other countries will also adopt similar processes, in the near future.
- Military systems and public order: Many countries also consider that AI, ML and ADM are significant for the control of public order through military systems and more is coming. 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.[13]

How might this breach equality and non-discrimination laws?

20. In summary, unlawful discrimination can occur when

- biased data sets are used to train a ML algorithm,
- algorithms are unlawfully indirectly discriminatory,

[13] 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

- AI based techniques are used as a form of harassment, or
- ADM discriminates on the grounds of protective characteristics.

21. The leaked EC document,[14] described some of the potential for discrimination arising from AI systems, in this way[15]–

> i. Risks for fundamental rights, including discrimination, privacy and data protection
>
> - Bias and discrimination arc inherent elements of any societal or economic activity. Human decision-making is also prone to mistakes and biases. However, the same level of bias when present in an artificial intelligence could affect and discriminate many people without the social control mechanisms that govern human behaviour. In addition to discrimination, artificial intelligence may lead to breaches of other fundamental rights,[16] including freedom of expression, freedom of assembly, human dignity, private life or right to a fair trial and effective remedy.
>
> - These risks might be a result of flawed design of artificial intelligence systems (e.g. the system is programmed to discard female job applications} or the input of biased data (e.g. the system is trained using only data from white males). They can also occur when the system 'learns' during the use phase, for example when an artificial intelligence system 'learns' that students with the best academic results share the same postal codes which happen to be prevalently white in population. The risks will in such cases not stem from a flaw in the original design of the system but from the practical impacts of the correlations or patterns that the system identifies in a large dataset.

22. This is not just a European problem; Regulators have already note it as occurring in the UK. For instance, the Financial Conduct Authority (FCA) identified that biased data sets were being used to calculate risk, and to set the price for insurance, in the household insurance market, finding[17] that–

[14] As I have noted, some caution must be exercised in considering this leaked draft for obvious reasons and there are real doubts that it is in fact very close to contemporary thinking in the EC

[15] Ibid.p.8

[16] Council of Europe research shows that a large number of fundamental rights could be impacted from the use of artificial intelligence: https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5

[17] See "Pricing practices in the retail general insurance sector: Household insurance", FCA Thematic Review TR18/4 October 2018, available at - https://www.fca.org.uk/publication/thematic-reviews/tr18-4.pdf

> 4.21… firms were using datasets (including datasets purchased from third parties) within their pricing models which may contain factors that could implicitly or potentially explicitly relate to race or ethnicity…
>
> 4.22 Firms were asked how they gained assurance that the third-party data they used in pricing did not discriminate against certain customers based on any of the protected characteristics under the Equality Act 2010. Many firms could not provide this assurance without first contacting the third-party provider. Further, some firms responded that they relied on the third-party provider to comply with the law and undertook no specific due diligence of their own to ensure that the data were appropriate to use.

23. I will look at some other examples in greater details later on in this paper.

Surely legislative steps already have already been taken to regulate AI systems to prevent discrimination, is that right?

24. In fact, no! This is exactly the point that the leaked draft made and it is widely acknowledged by others. So far, no legislation has been passed that has been designed specifically to tackle discrimination in AI systems. There has been a great deal of discussion about the ethics of creating AI systems, and this has almost always noted the potential for discrimination. Indeed, this discussion has been a spur to consideration as to whether there is a need for specific legislation.

25. One view, though, is that we already have some good tools to address discriminatory AI. As I shall explain in a moment there is some truth in this but there are real problems in really getting to grips with discriminatory AI with the legal tools that we currently have. However, before we look at how the Equality Act 2010 and other legislation can help, it is important to be aware of what are the commonly accepted ethical principles that apply to the development of AI systems. Understanding these is essential when analysing how potentially indirectly discriminatory AI systems might be justified.

26. This discussion of the ethics of AI has gone on for some time but it has accelerated in the course of 2019. It has largely been undertaken by the business and engineering community with some input from philosophers and others involved in social sciences. Rather less input has been provided by lawyers.

27. Nonetheless there is an emerging consensus across much of the world. A key player has been the Institute of Electrical and Electronics Engineers[18] (IEEE[19]) which published its 'Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems' on the 25th March 2019 (EAD1e). We all need to be aware of them!

28. EAD1e states that the development, and implementation of AI/IS should have an ethical and values-based design, guided by the following eight General Principles –

> 1. Human Rights A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.
>
> 2. Well-being A/IS creators shall adopt increased human well-being as a primary success criterion for development.
>
> 3. Data Agency A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.
>
> 4. Effectiveness A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.
>
> 5. Transparency The basis of a particular A/IS decision should always be discoverable.
>
> 6. Accountability A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.
>
> 7. Awareness of Misuse A/IS creators shall guard against all potential misuses and risks of A/IS in operation.
>
> 8. Competence A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

29. EAD1e also states that there must be legal frameworks for accountability, stating –

> • Autonomous and intelligent technical systems should be subject to the applicable regimes of property law.

[18] This is one of the world's leading consensus building organizations aiming to nurture, develop and advance global technologies.

[19] Referred to as "eye triple e".

• Government and industry stakeholders should identify the types of decisions and operations that should never be delegated to such systems. These stakeholders should adopt rules and standards that ensure effective human control over those decisions and how to allocate legal responsibility for harm caused by them.

• The manifestations generated by autonomous and intelligent technical systems should, in general, be protected under national and international laws.

• Standards of transparency, competence, accountability, and evidence of effectiveness should govern the development of autonomous and intelligent systems.

30. The EAD1e principles have been hugely important in the development of other thinking. They have been largely adopted by the OECD on the 22nd May 2019[20] and then in turn by G20 group of countries on the on the 8th and 9th June 2019.[21] Meanwhile the AI HLEG had been working away, and its proposed principles were also finalized in June 2019. These can be found here[22] and are again a "must – read" for anyone involved with these issues. The key points are these –

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.

[20] See https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[21] See https://www.mofa.go.jp/files/000486596.pdf

[22] See https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

> 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.
>
> Societal and environmental well-being: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes.

31. Overall there is an emphasis that AI systems must be "human-centric". There is obviously much overlap between the AI HLEG principles and EADIe. It is expected that any proposal for legislation brought forward in the first half of 2020 by the European Commission will reflect all these principles in some form.

But, what legal tools to address discriminatory AI systems do we *currently* have?

32. Ethical principles are not enough to stop discrimination of course, so it is important to know what tools can already be brought to bear on these problems. The answer is that the general laws that we have are the Equality Act 2010, the Human Rights Act 1998, the General Data Protection Regulation (GDPR)[23] and the Law Enforcement Directive (LED).[24]

Equality legislation

33. We all know that our Equality Act 2010 is designed to protect against direct and indirect discrimination, harassment, and disability related discrimination, in relation to the defined list of protected characteristics. It is good that the sphere of application of the Equality Act 2010 does include all the main areas in which AI systems are being deployed.

34. The Equality Act 2010 is assisted by the Human Rights Act 1998 which gives effect to Article 14 of the European Convention on Human Rights (ECHR). The HRA protects against discrimination on the protected grounds and beyond in

[23] 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.

[24] 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.

contexts where the other main protections of the ECHR such as the right to private life in Article 8 are engaged. You will be broadly familiar with these protections so I will not elaborate them here.

The Black box problem

35. There is one major problem though in using equality laws such as these. Although the ethical principles set out above constantly emphasise the need for transparency and explainability in the use of AI, yet this is commonly not the case. Or rather there are often disputes as to how much has to be explained and made transparent.

36. Through the work of the AI Law Consultancy Dee Masters and I are convinced that many organisations will struggle to demonstrate that their AI systems are transparent due to what has come to be called the "black box" problem. This problem arises because in many cases it is impossible to peer inside an algorithm, AI or ML process, so as to understand how decisions are being made.

37. The AI HLEG noted this problem, saying[25] –

> **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.

38. Jenny Burrell, an academic who specialises in AI, put the point in this way[26] -

> 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).

39. When an AI system is not sufficiently transparent and leads apparently to *prima facie* discrimination, we can foresee several difficulties. There will be issues for

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

[26] Jenna Burrell, "*How the machine 'thinks': Understanding opacity in machine learning algorithms*", 2016: https://journals.sagepub.com/doi/10.1177/2053951715622512

the complainant of course but also we predict that organisations will encounter great difficulties in proving objective justification.

40. Indeed, it is very possible that the lack of transparency itself will lead to the courts finding that the technology is *prima facie* discriminatory. 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 C-109/88 *Danfoss*[27] and has been reiterated on many occasions. We see 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, to succeed.

Data Protection Legislation (GDPR and LED)

41. The GDPR can help with this. Many equality and employment lawyers do not know as much about the GDPR and the LED as they perhaps should, if they are to deal with these issues. The EU has extensive protections against the misuse of data, and these are important also in preventing discriminatory outcomes, because all AI, ML and ADM systems involve the processing of data.

42. The starting point is Article 8 of the Charter of Fundamental Rights of the European Union (CFREU) which 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.

43. The GDPR, together with the LED, were designed to give more detailed protection of natural persons with regard to the processing of personal data (Article 1). 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 –

[27] Case C- 109/88, *Handels- og Kontorfunktionaerernes Forbund i Danmark v Dansk Arbejdsgiverforening Ex p. Danfoss A/S*

> 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.

44. Article 6(1)(e) and (f) GDPR 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.

45. 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 algorithm (whether or not underpinned by machine learning) where there are legal consequences for him or her or similarly significant repercussions including decisions that are discriminatory. Article 22 says as follows -

> 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.
>
> Paragraph 1 shall not apply if the decision:
>
> is necessary for entering into, or performance of, a contract between the data subject and a data controller;
>
> 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
>
> is based on the data subject's explicit consent.

46. However, like Article 21, the right created by Article 22 is limited in many ways as will be seen if the full text is studied.

47. The LED[28] covers 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, 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.

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

Using the GDPR to open the black box

49. Importantly for our discussion of how to deal with the black box problem, the GDPR specifically refer to the principle of transparency; there is now an important debate in Europe as to the extent to which these principles might be used to force organisations to disclose the contents of their "black box". The current position of the European Data Protection Board (EDPB) is that the GDPR does *not* go so far as to dictate that algorithms or the basis for ML must be completely disclosed.[29] It considers simply 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.

50. So, while the GDPR may be useful in terms of seeking to peer inside the "*black box*" it is currently thought that it is unlikely to compel complete transparency. There is a real concern as to whether this is enough. There is however an important consultation on foot about this issue being driven by the Information

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

[29] https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

Commissioner's Office (ICO).[30] Although this is likely to be closed by the time that this lecture is given it is probably worth considering it and seeing what the ICO comes up with in due course.

Some examples of discriminatory AI systems

51. In this next section I shall discuss further some typical examples of discriminatory AI systems, in order to try to convey more fully the extent of the issues that are arising.

A scary story from the US concerning facial recognition technology (FRT)

52. A good place to start is with a scary story from the United States (US) where a recruitment firm called HireVue (which has also launched in Europe) has a huge following, as its website says[31] –

> HireVue, recruiters and hiring managers make better hiring decisions, faster. HireVue customers decrease time to fill up to 90%, increase quality of hire up to 88%, and increase new hire diversity up to 55%.
>
> HireVue is the market leader in video interviewing and AI-driven video and game-based assessments.
>
> HireVue is available in over 30 languages and has hosted more than 10 million on-demand interviews and 1 million assessments.
>
> Our more than 700 customers worldwide include over one-third of the Fortune 100 and leading brands such as Unilever, JP Morgan Chase, Delta Air Lines, Vodafone, Carnival Cruise Lines, and Goldman Sachs. Learn more about HireVue's approach

53. It is widely thought that HireVue's FRT approach to determining who would be truthful loyal good fit employees discriminates heavily against those with the protected characteristic of disability. Scholars (principally from the New York bases AI Now Institute) commented in November 2019[32] -

[30] See https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-and-the-turing-consultation-on-explaining-ai-decisions-guidance/

[31] https://www.hirevue.com/why-hirevue

[32] Meredith Whittaker, AI Now Institute at NYU et al., Disability, Bias, and AI, November 2019, see https://ainowinstitute.org/disabilitybiasai-2019.pdf

> The norms enforced and encoded by AI systems have significant material consequences. Artificial intelligence is rapidly being introduced into the workplace, and is already informing decisions about hiring, management, performance assessment, and beyond.[56]
>
> The example of the AI company HireVue is instructive. The company sells AI video-interviewing systems to large firms, marketing these systems as capable of determining which job candidates will be successful workers, and which won't, based on a remote video interview. HireVue uses AI to analyze these videos, examining speech patterns, tone of voice, facial movements, and other indicators. Based on these factors, in combination with other assessments, the system makes recommendations about who should be scheduled for a follow-up interview, and who should not get the job.[57]
>
> In a report examining HireVue and similar tools, authors Jim Fruchterman and Joan Mellea are blunt about the way in which HireVue centers non-disabled people as the "norm," and the implications for disabled people: "[HireVue's] method massively discriminates against many people with disabilities that significantly affect facial expression and voice: disabilities such as deafness, blindness, speech disorders, and surviving a stroke."[58]

Footnotes

56 See, for example, Jaden Urbi, "Some Transgender Drivers Are Being Kicked Off Uber's App," CNBC, August 8, 2018, https://www.cnbc.com/2018/08/08/transgender-uber-driver-suspended-tech-oversight-facial-recognition.h tml; and Sarah Kurbi, "The New Ways Your Boss Is Spying on You," Wall Street Journal, July 19, 2019, https://www.wsj.com/articles/the-new-ways-your-boss-is-spying-on-you-11563528604.

57 HireVue website, HireVue, accessed October 23, 2019, https://www.hirevue.com/ .

58 Jim Fruchterman and Joan Mellea, "Expanding Employment Success for People with Disabilities," Benetech (November 2018), https://benetech.org/wp-content/uploads/2018/11/Tech-and-Disability-Employment-Report-November-20 18.pdf .

54. HireVue's approach may discriminate on other grounds and it is good to know that there is widespread skepticism as to its efficacy.[33] There has been some pushback, as you will find on www.ai-lawhub.com; for instance the State of Illinois has passed an Artificial Intelligence Video Interview Act,[34] in which the key provision section 5 states that –

> An employer that asks applicants to record video interviews and uses an artificial intelligence analysis of the applicant-submitted videos shall do all of the following when considering applicants for positions based in Illinois before asking applicants to submit video interview

[33] See e.g. Manish Raghavan, Solon Barocas, Jon Kleinberg, Karen Levy, Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices, Cornell University v.3 6 Dec 2019, see https://arxiv.org/abs/1906.09208

[34] http://www.ilga.gov/legislation/publicacts/fulltext.asp?Name=101-0260

> (1) Notify each applicant before the interview that artificial intelligence may be used to analyze the applicant's video interview and consider the applicant's fitness for the position
>
> (2) Provide each applicant with information before the interview explaining how the artificial intelligence works and what general types of characteristics it uses to evaluate applicants
>
> (3) Obtain, before the interview, consent from the applicant to be evaluated by the artificial intelligence program as described in the information provided.
>
> An employer may not use artificial intelligence to evaluate applicants who have not consented to the use of artificial intelligence analysis.

55. Possible proposals in the leaked EC's draft document[35] seem to parallel these concerns as it is reported to say[36] that the –

> …use of facial recognition technology by private or public actors in public spaces would be prohibited for a definite period (e.g. 3–5 years) during which a sound methodology for assessing the impacts of this technology and possible risk management measures could be identified and developed.

Is FRT really that scary? Why should we worry?

56. As humans we 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. FRT is becoming increasingly cheap to purchase and its deployment is increasing all the time.

[35] https://www.euractiv.com/section/digital/news/leak-commission-considers-facial-recognition-ban-in-ai-white-paper/

[36] Though there is also some scepticism that the EC will formally propose a ban on FRT.

57. Many well–known companies such as Amazon,[37] IBM,[38] and Microsoft[39] have relatively cheap proprietary FRT products; in some circumstances access to these is free. Other companies offer dedicated FRT products promising very high level of utility in specific circumstances. So, it is increasingly being used to verify identities as a gateway to access or deny access to a range of goods facilities and services. Last year for instance the London Evening Standard said[40] not entirely fancifully that "Facial recognition [would] tell pub staff who's next in queue for a pint." It is being used by some clubs to identify drug dealers and other ne'er-do-wells so that they can be evicted or prevented from admittance.[41]

58. The problem is that it is very well established that FRT systems will provide false matches or sometimes fail to make matches when they would be appropriate. These are false positives and false negatives and 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 data bases 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.[42] 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

---

[37] For instance see https://aws.amazon.com/free/machine-learning/?trk=ps_a131L0000057ji8QAA&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

[38] For instance see https://cloud.ibm.com/catalog/services/visual-recognition

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

[40] https://www.standard.co.uk/tech/facial-recognition-to-tell-pub-staff-whos-next-in-queue-for-pint-a4203011.html

[41] https://www.theguardian.com/technology/2019/oct/05/facial-recognition-technology-hurtling-towards-surveillance-state

[42] Joy Buolamwini, Timnit Gebru; Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91, 2018, "*Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*":see http://proceedings.mlr.press/v81/buolamwini18a.html

were to be used in Europe as a gateway to a benefit or service of some kind it would be potentially discriminatory.

59. 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. [43] It noted that –

> … 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.

60. 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[44] –

> …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.

[43] Patrick Grother, Mei Ngan, Kayee Hanaoka, "Face Recognition Vendor Test (FRVT) part 3: Demographic Effects", see https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

[44] Ibid.

Can FRT perhaps be justified?

61. In such a scenario, there will be unlawful indirect 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[45] the jurisprudence of the ECtHR will also entail asking whether this use is appropriate and necessary within a democratic context.

62. What is involved? 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.[46]

63. 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 Dee Masters and I can see that 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 if it produced numerous false positive and numerous false negatives. It may have an even bigger problem in showing that the aim was being achieved in a proportionate way. We discuss this in some detail on www.ai-lawhub.com , see in particular here.[47]

64. This is for two reasons –

    a. As we have noted, much research which 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

[45] This is likely also to be the case if the user is a private company too.

[46] C-17084 *Bilka-Kaufhaus GmbH v Weber von Hartz*

[47] https://ai-lawhub.com/government/#a

individuals[48]. 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. We consider that any justification defence in relation to the use of this system would fail because it can hardly be said that its aim is being achieved.

b. 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 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.

65. When it comes to deciding whether the aim of using FRT is legitimate the ethical principles identified above will be very important. The over-arching theme is sometimes said to be that AI systems must be "human-centric", and this is to be achieved by working to the standards set by the AI HLEG. [49]

[48] Professor Pete Fussey & Dr Daragh Murray, "*Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology*", July 2019: see https://48ba3m4eh2bf2sksp43rq8kk-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report.pdf

[49] The administrative court has begun to consider some of these issues in *Bridges, R (On Application of) v The Chief Constable of South Wales Police* [2019] WLR(D) 496, [2019] EWHC 2341 (Admin); see https://www.bailii.org/ew/cases/EWHC/Admin/2019/2341.html It is understood that an application to appeal this judgment has been made.

66. So, an aim will only be legitimate for the purpose of an objective justification defence in so far as the AI system is intended to achieve an aim consistent with these principles. Accordingly, we consider it very likely that FRT that gives rise to *prima facie* indirect discrimination, will only be justifiable in so far as it also does not undermine 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, the use of facial recognition in more mundane commercial contexts may well be incapable of justification if the law develops in the direction which we anticipate.

67. The paper, "*The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*" will be highly relevant to the question of proportionality. Perhaps in the end the failure to open the black box will also be fatal.

ADM that discriminates by drawing sexist conclusions

68. A different example will point up how badly thought-through AI systems can cause direct discrimination. It concerns Louise Selby, who is a medical doctor. She was also a client of PureGym the well-known commercial gym company.

69. A problem occurred when she was unable to use a swipe card provided by the company to access locker rooms at one of its venues. When the problem was investigated it transpired that the company was using a computer system that used a member's title to determine which changing room (male or female) a customer would be permitted to access.[50] It was all done by ADM in its AI system.

70. The computer system had an algorithm that searched the data base of the gym company's members, to identify their gender and then allocated 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.

71. This was a classic case of direct sex discrimination. The customer was treated less favourably because she was a woman in circumstances in which a

[50] https://www.informationsociety.co.uk/pure-gym-in-cambridge-sexist-computer-assumed-thiswoman-dr-louise-selby-was-a-man-because-she-is-a-doctor

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 forthwith that it had made 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.

Indirect discrimination in risk assessments

72. Many public authorities are using AI systems to predict the risk of a certain occurrence for instance -

- 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.

73. Risk analysis is a key area where discrimination can occur in a way which can have significant effects on individuals, as the FCA pointed out. To exemplify this, Dee Masters and I have analysed the use of "Risk-Based Verification" (RBV) systems used by local authorities to predict the risk of the misallocation of Housing Benefits and Council Tax Benefits. The law imposes no fixed verification process but local authorities can ask for documentation and information from any applicant "*as may reasonably be required*".[51] 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.[52]

74. The 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*".[53] For example, an individual with a low risk might simply need to provide proof of

[51] The Council Tax Benefit Regulations 2006, SI 2006 No. 215, reg 72: 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

[52] Housing Benefit and Council Tax Benefit Circular, HB/CTB S11/2011:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633018/s11-2011.pdf

[53] Ibid.

identity but someone with a high-risk rating might be subject to Credit Reference Agency checks, visits, increased documentation requirements etc.[54]

75. A DWP circular shows that the Department is aware that ML algorithms are being deployed as part of this process. [55] 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 instance, 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.[56]

76. 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 United Kingdom'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.[57]

77. Accordingly, it is possible that these RBV systems utilised across the UK the United Kingdom could be acting in a discriminatory way. However, because of the "*black box*" problem we have 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 discrimination free. Accordingly, we anticipate that AI systems which predict risk, but which 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.

78. We have discussed the similar problems associated with RBV in the context of settled status on www.ai-lawhub.com

[54] Ibid.

[55] Ibid.

[56] Ibid.

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

Harassing Deepfakes

79. First what are Deepfakes? This is how the BBC explains them[58] –

> If you are active on social media, you might have seen many apps and filters used to swap faces in pictures and videos. That technology has been around for many years but has rarely produced believable results.
>
> In 2017, a few posts emerged on Reddit, showing how it was possible to use Artificial Intelligence to seamlessly swap faces in videos. That technique was called deepfake. The 'deep' bit comes from Deep Learning, a branch of AI that uses something known as neural networks. In a nutshell, neural networks are a type of machine learning technique that bears some resemblance to how the human brain works.
>
> Nowadays, there are several different ways to swap faces in a very realistic way. Not all use AI, but some do: deepfake is one of them.
>
> The term deepfake is now generically used by the media to refer to any video in which faces have been either swapped or digitally altered, with the help of AI ...

80. It is obvious that these can be used to harass people, both contrary to the Prevention of Harassment Act 1997 and the EA. At first celebrities were most affected and many complained about their use,[59] but now the technology is available to be used quite freely so that the BBC has even discussed how to go about making a Deepfake.[60] It is obvious that they can be used for revenge porn videos or in similar ways to harass quite ordinary non-celebrity individuals. The potential to make Deepfakes is growing as an issue very fast. The American Bar Association kicked off 2020 with a blog on the 1st January entitled "Detecting Deepfakes" which showed how Deepfake videos (and indeed audios) are becoming harder to identify,[61] and the issue is now recognised as being so serious that on the 6th January 2020, Monika Bickert,

[58] https://www.bbc.co.uk/bitesize/articles/zfkwcqt

[59] See for instance, Danny Fortson, The rise of deepfakes: what are they and how can we know what's real?, Sunday Times 22 December 2019; see https://www.thetimes.co.uk/article/the-rise-of-deepfakes-what-are-they-and-how-can-we-know-whats-real-x03jp3rqr

[60] Ibid. It suggests that this could be done for fun; it is not suggested that the BBC encourages harassment!

[61] Sharon D. Nelson, John W. Simek and Michael Maschke, Detecting Deepfakes: Deepfake videos are becoming harder to identify and may threaten the 2020 election; see https://www.americanbar.org/groups/law_practice/publications/law_practice_magazine/2020/jf2020/jf20nelsonsimekmaschke/

Vice President, Global Policy Management at Facebook and Instagram, recognising this problem said that Deepfakes would be banned,[62] –

> Going forward, we will remove misleading manipulated media if it meets the following criteria:
>
> - It has been edited or synthesized – beyond adjustments for clarity or quality – in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say. And:
> - It is the product of artificial intelligence or machine learning that merges, replaces or superimposes content onto a video, making it appear to be authentic.
>
> This policy does not extend to content that is parody or satire, or video that has been edited solely to omit or change the order of words.
>
> Consistent with our existing policies, audio, photos or videos, whether a deepfake or not, will be removed from Facebook if they violate any of our other Community Standards including those governing nudity, graphic violence, voter suppression and hate speech…

81. This does not mean that Deepfakes will stop. The only safe assumption is that they will not; we shall see them appear on other platforms and even still on Facebook and Instagram. They are though simply a form of harassment which can be made subject to the EA, or the HRA, or in some circumstances the could be subject to defamation laws.

What's going to happen to address these problems and make them easier to resolve and prevent discrimination occurring?

82. At present many of those working as technicians and scientists in the field of AI systems are aware of these problems. They are also aware that there are real dangers ahead in machines taking over without humans understanding what is happening or how it has occurred. This has led to a worldwide discussion of the right ethical approach to AI systems.

83. Within the UK we have our own Centre for Data Ethics and Innovation (CDEI)[63] which has undertaken some work in this field publishing papers on Deepfakes, SmartSpeakers and Voice Assistants, and Personal Insurance. In

[62] https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/
[63] See https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation

respect of personal insurance, they were fully aware of the problem of ADM setting insurance pricing in a discriminatory way[64]–

> Insurers are prohibited by law from basing pricing and claims decisions on certain protected characteristics, including sex and ethnicity. However, other data points could feasibly act as proxies for these traits, for example with postcodes signalling ethnicity or occupation categories signalling gender. This means that AI systems can still be trained on datasets that reflect historic discrimination, which would lead those systems to repeat and entrench biased decision making.

84. Indeed there is a lot of work on AI going on here from other bodies,[65] and there are other specific regulators who will have to address these issues within their own domains; these include the Surveillance Camera Commissioner, the Biometrics Commissioner and the National Data Guardian for Health and Social Care.

85. However, we think it is critical that the Equality Commissions get their act together on these issues because at present neither addresses AI systems issues in their workplans or on their website at all as far as we can see. We believe that the European association of Equality Bodies – Equinet, which we have been advising can help 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.[66]

86. It is really important that more *European* and *UK* based work is done on the implications for equality law. At present so much work concerning the development of AI systems is taking place in the United States of America (US)

---

[64] See https://www.gov.uk/government/publications/cdei-publishes-its-first-series-of-three-snapshot-papers-ethical-issues-in-ai/snapshot-paper-ai-and-personal-insurance

[65] Such as such as: the AI Council; the Office for AI; the House of Lords Select Committee on AI; the House of Commons Inquiry on Algorithms in Decision-Making; the Alan Turing Institute; the National Data Guardian; the Information Commissioner's Office; a proposed new digital regulator; departmental directorates; the Office for Tackling Injustices; the Regulatory Horizons Council; Ofcom;1 NHSX, a new health National AI Lab; AI monitoring by the Health and Safety Executive's Foresight Centre; AI analysis from the Government Office for Science; the Office for National Statistics' Data Science Campus; and the Department of Health and Social Care's code of conduct for data-driven health and care technology.

[66] This paper does not discuss the role that AI systems can have in analysing data bases for bias. Nonetheless it is worth recognising that this is one 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" (see http://aif360.mybluemix.net/?utm_campaign=the_algorithm.unpaid.engagement&utm_source=hs_email&utm_medium=email&utm_content=69523284&_hsenc=p2ANqtz-9vaujms_IQeQkh8nE92xGK7pisSc5eYX3nQkytSKQkCd7rAAd2pPmn_kgregFKWVMMD7G0LuVo_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, facial recognition technology.

or is backed by capital sourced through the US that the debate has a trans-Atlantic slant, perhaps because 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 challenges of that sort. While a certain amount of work is being done there in relation to the interface between programming for AI systems and the need to avoid discrimination, this is proceeding on the basis of the US concepts of equality which do not wholly coincide with EU and UK concepts. [67]

The Council of Europe

87. As well as within the European Commission and the AI HLEG, much important work is also being done by the Council of Europe which is responsible for the ECHR. Following Brexit this may be most important for the UK if it is really not to be a rule-taker from Europe. In our view the CoE's programme of work in this field is already well-developed and should be actively monitored.

88. In 2018, the CoE published an excellent standard-setting document entitled "*Discrimination, AI, and algorithmic decision-making*" written by Prof. Frederik Zuiderveen Borgesius Professor of Law.[68] Since then, the CoE has developed a website dedicated to addressing human rights issues raised by artificial intelligence.[69] 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). We understand that CAHAI will 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 AI, based on CoE's standards on human rights, democracy and the rule of law. Marija Pejčinović Burić, Secretary General of the Council of Europe, 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."

[67] Of course, it is important that such work should be done, but it must be noted, for instance, that the approach to indirect discrimination is significantly different: see e.g. Manish Raghavan, Solon Barocas, Jon Kleinberg, Karen Levy, ibid. at p. 4 "A brief overview of U.S. employment discrimination."

[68] https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

[69] https://www.coe.int/en/web/artificial-intelligence/home

89. 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).[70]

90. Most recently, in 2019, it produced a practical guide called "*Unboxing Artificial Intelligence: 10 steps to protect human rights*".[71]

The Fundamental Rights Agency

91. Another important player is the EU's Fundamental Rights Agency (FRA). 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.[72]

92. 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.[73]

93. 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. [74]

94. 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.[75]

[70] https://www.coe.int/en/web/freedom-expression/msi-aut

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

[72] https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf

[73] https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-preventing-unlawful-profiling-guide_en.pdf

[74] https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf

[75] https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper.pdf

95. The FRA has also collated a very detailed record of the resources currently available.[76] This is a key resource.

Conclusions

96. We hope that you too will play your part in encouraging the use of non – discriminatory AI systems by business and public authorities and by challenging systems that do not comply with the ethical principles set out above. There is no doubt that at present there is a will to avoid discrimination in Europe but we well know that it can also be used as a very powerful force to deny human rights and equality as well so understanding what can happen and what is available to challenge it is critical. Please keep in touch with the consultancy as set out on the first page of this paper.

ROBIN ALLEN QC

20th January 2020
© 2020 All rights reserved

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

Cloisters is at the heart of all the major equality and discrimination cases and our barristers play an important role in the development of law and policy in this area. We cover every aspect of discrimination and equality law and work on the most complex, high profile and high value cases as advisers and advocates.

Cloisters were once again ranked a leading set in Chambers and Partners 2020 and Legal 500 2020, and won Chambers and Partners prestigious Employment Set of the Year Award 2018.

> *"Simply the go-to set if you want to win an Equality Act case. It fields counsel at all ranges of the spectrum, all of whom are accessible, down to earth and able to put clients at ease."*
>
> (Legal 500)

OUR CLIENTS

We represent employers and employees in multinationals, SMEs, non-government organisations, charities, government departments, regulators, local authorities, associations and individuals. Many of our members are qualified Public Access barristers especially trained to provide legal advice and litigation support directly to organisations.

OUR TRAINING SERVICE

Cloisters offers CPD accredited training courses, workshops and seminar programmes. Cloisters training programmes are highly regarded and well attended. Our members can deliver bespoke in-house training to law firms and professional bodies at their premises.

OUR PEOPLE

We have a well-earned reputation for fearless advocacy and the highest of standards. Over three quarters of our employment team are ranked leaders. Many of our barristers achieve ground-breaking results and appear in almost all major employment litigation and landmark cases.

CLOISTERS' EMPLOYMENT AND DISCRIMINATION BARRISTERS

Robin Allen QC (1995)
Jonathan Mitchell QC (1992)
Brian Napier QC (2002)
Paul Epstein QC (2006)
Daphne Romney QC (2009)
Jacques Algazy QC (2012)
Caspar Glyn QC (2012)
Jason Galbraith-Marten QC (2014)
Rachel Crasnow QC (2015)
Schona Jolly QC (2017)
Tom Coghlin QC (1918)
Ed Williams QC (2018)
Andrew Buchan (1981)
Declan O' Dempsey (1987)
Michael Potter (1988)

Catherine Casserley (1991)
Yvette Genn (1991)
Paul Michell (1991)
John Horan (1993)
Sally Cowen (1995)
Sally Robertson (1995)
David Massarella (1999)
Akua Reindorf (1999)
Tom Brown (2000)
Claire McCann (2000)
Anna Beale (2001)
Adam Ohringer (2001)
Dee Masters (2004)
Sarah Fraser Butlin (2005)
Daniel Dyal (2006)
Chris Milsom (2006)

Olivia-Faith Dobbie (2007)
Will Dobson (2008)
Caroline Musgrave (2008)
Catriona Stirling (2008)
Nathaniel Caiden (2009)
Sheryn Omeri (2010)
Chesca Lord (2011)
Rachel Barrett (2012)
Jennifer Danvers (2012)
Tamar Burton (2012)
Tom Gillie (2013)
Navid Pourghazi (2014)
Jonathan Cook (2017)
Ruaraidh Fitzpatrick (2017)
Catherine Meenan (2018)