

RESPONSE TO THE EUROPEAN COMMISSION
WHITE PAPER “ON ARTIFICIAL INTELLIGENCE - A EUROPEAN
APPROACH TO EXCELLENCE AND TRUST”

Robin Allen QC and Dee Masters, barristers,
AI Law Consultancy (www.ai-lawhub.com) and Cloisters chambers
(www.cloisters.com)

Cloisters, 1 Pump Court, Temple, London, EC4Y 7AA

ra@cloisters.com and deemasters@cloisters.com

Contents

<i>INTRODUCTION</i>	3
<i>OUR EXPERTISE</i>	3
<i>SUMMARY OF SUBMISSIONS</i>	3
<i>EXTENDED ANALYSIS</i>	6
1. There is currently widespread mistrust of AI in society as a whole. This view has been formed by the public despite an incomplete understanding of the risk which AI poses to the principle of non-discrimination. We expect that as society better understands the many ways in which AI can discriminate against individuals, there will be far higher levels of mistrust.....	6
2. We strongly welcome the Europe Commission’s proposal to regulate AI, with a particular focus on equality, so as to create “ <i>trust</i> ” from the public and businesses alike.....	7
3. Since AI can discriminate in multiple ways, the first step must be to ensure that equality law within Europe covers all sectors (employment, goods, facilities and services) and each of the protected characteristics (sex, age, disability, race, sexual orientation etc).....	7
4. To support a universal principle of non-discrimination, there is merit in introducing targeted procedurally based safeguards with the intention that these will encourage the development and use of systems which comply with the principle of non-discrimination.	8
5. Procedurally based rules which should be considered as a means of encouraging businesses to comply with the principle of equality are as follows: an AI register (for example, those with “high risk” applications of AI), mandatory AI auditing, a requirement to publish audit documentation and specialised procedural rules concerning the processing of biometric data.	9
6. There is merit in limiting new procedurally based safeguards to “ <i>high risk</i> ” applications of AI only.....	11

7. AI should be classed as high risk where it “*produces legal effects*” for individuals or “*similarly significant effects*” so as to dovetail with Article 22 of the GDPR..... 11

8. The European Commission should create or support a programme of inquiry so as to ensure that the relevant “*high risk*” applications of AI are identified..... 12

9. The introduction of targeted procedural rules is very different to the use of *standards*. The principle of non-discrimination is universal, and the substantive requirements of equality should never be targeted at particular products or sectors using a risk-based approach..... 13

10. A certification scheme should be introduced for “*high risk*” applications which indicate when AI has features that are consistent with the principle of non-discrimination, for example, it utilises a balanced data set or it is “*human centric*” 14

11. There should be a prohibition on decisions being taken by solely an AI system in a way that mirrors Article 22 of the GDPR. 14

12. All organisations involved in the development of AI should be liable for any discrimination including the company, organisation or public body that ultimately uses it. 14

13. To reflect the difficulties that some end users will face determining whether the AI system is non-discriminatory, due to the “*black box*” problem, end users should be able to rely on a defence that they took all reasonable steps to ensure that the AI system was non-discriminatory. However, this defence should not be available to those who had manufactured or supplied such systems to end-users since they are in a position to ensure that the AI is non-discriminatory..... 16

14. To support a universal principle of non-discrimination, the burden of proof should shift to the Defendant where there is a lack of transparency and some evidence to suggest that discrimination is occurring.
17

15. The European Union should take steps to ensure that the enforcement of the right to equal treatment (the principle of non-discrimination) and the protection of data rights are fully aligned..... 17

16. The European Union must not permit international trade rules to be developed that in any way undermine the right to equality by immunising intellectual property rights from disclosure when necessary and appropriate for the enforcement of those equality rights. 18

Introduction

In this paper we set out 16 overarching comments in response to the European Commission papers “*On Artificial Intelligence – A European approach to excellence and trust*”¹ (**the First White Paper**) and “*Commission’s Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and Robotics*”² (**the Second White Paper**). We focus solely on the proposals to regulate AI so as to effectively enshrine the principle of non-discrimination.

Our expertise

We are specialist discrimination law barristers based in London working from Cloisters chambers.³ We have a strong interest in the interplay between Artificial Intelligence (AI) and discrimination law. We work together as the AI Law Consultancy and maintain an online resource dedicated to AI, human rights, discrimination and data protection.⁴ Our recent work includes writing Equinet’s recent paper “*Regulating for an equal AI: Equality bodies working in partnership for a new European approach to equality and Artificial Intelligence: meeting the new challenges to equality and non-discrimination from increased digitalisation and the use of Artificial Intelligence*” (**the Equinet Report**) released on 10 June 2020.⁵ Our report was anticipated by the First White Paper.⁶

We have been consulted by the UK’s Centre for Data Ethics & Innovation (CDEI), which is an advisory body set up by the UK Government to seek to maximise the benefits of data-enabled technologies, including artificial intelligence,⁷ and we have been retained by the UK Trade Union Congress (TUC) to advise in relation to the employment implications of AI and other technologies in the light of COVID-19.

Summary of submissions

1. There is currently widespread mistrust of AI in society as a whole. This view has been formed by the public despite an incomplete understanding of the risk which

¹ https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

² https://ec.europa.eu/info/files/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics_en

³ www.cloisters.com

⁴ www.ai-lawhub.com

⁵ https://equineteurope.org/wp-content/uploads/2020/06/ai_report_digital.pdf

⁶ See the First White Paper at fn 35.

⁷ <https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation>

AI poses to the principle of non-discrimination. We expect that as society better understands the many ways in which AI can discriminate against individuals there will be far higher levels of mistrust.

2. We strongly welcome the Europe Commission's proposal to regulate AI, with a particular focus on equality, so as to create "*trust*" from the public and businesses alike.
3. Since AI can discriminate in multiple ways, the first step must be to ensure that equality law within Europe covers all sectors (employment, goods, facilities and services) and each of the protected characteristics (sex, age, disability, race, sexual orientation etc).
4. To support a universal principle of non-discrimination, there is merit in introducing targeted procedurally based safeguards with the intention that these will encourage the development and use of systems which comply with the principle of non-discrimination.
5. Procedurally based rules which should be considered as a means of encouraging businesses to comply with the principle of equality are as follows: a register of all significant uses of AI systems (for example, those with "*high risk*" applications of AI), mandatory AI auditing, a requirement to publish audit documentation and specialised procedural rules concerning the processing of biometric data.
6. There is merit in limiting new procedurally based safeguards to "*high risk*" applications of AI only.
7. AI should be classed as high risk where it "*produces legal effects*" for individuals or "*similarly significant effects*" so as to dovetail with Article 22 of the GDPR.
8. The European Commission should create or support a programme of inquiry so as to ensure that the relevant "*high risk*" applications of AI are identified.
9. The introduction of targeted procedural rules is very different to standards setting. The principle of non-discrimination is universal and the substantive requirements of equality should never be targeted at particular products or sectors using a risk-based approach.
10. A certification scheme should be introduced for "*high risk*" applications which indicate when AI has features that are consistent with the principle of non-discrimination, for example, it utilises a balanced data set or it is "*human centric*".

11. There should be a prohibition on decisions being taken solely by an AI system in a way that mirrors Article 22 of the GDPR.
12. All organisations involved in the development of AI should be liable for any discrimination including the company, organisation or public body that ultimately uses it.
13. To reflect the difficulties that some end users will face determining whether an AI system is non-discriminatory due to the “*black box*” problem, end users should be able to rely on a defence that they took all reasonable steps to ensure that the AI system was non-discriminatory. However, this defence should not be available to those who had manufactured or supplied such systems to end-users since they are in a position to ensure that the AI is non-discriminatory.
14. To support a universal principle of non-discrimination, the burden of proof should shift to the Defendant where there is a lack of transparency and some evidence to suggest that discrimination could be occurring
15. The European Union should take steps to ensure that the enforcement of the right to equal treatment (the principle of non-discrimination) and the protection of data rights are fully aligned.
16. The European Union must not permit international trade rules to be developed that in any way undermine the right to equality by immunising intellectual property rights from disclosure when necessary and appropriate for the enforcement of those equality rights.

Extended Analysis

1. There is currently widespread mistrust of AI in society as a whole. This view has been formed by the public despite an incomplete understanding of the risk which AI poses to the principle of non-discrimination. We expect that as society better understands the many ways in which AI can discriminate against individuals, there will be far higher levels of mistrust.

The First White Paper correctly identifies that AI has the potential to create enormous improvements within society but for that to happen effectively there must be guarantees that it is “trustworthy”.⁸

At present, there is a significant “trust” deficit. A recent paper in the UK by doteveryone called “People, Power and Technology: The 2020 Digital Attitudes Report” illustrates the extent of the problem⁹:

- 58% of survey respondents were concerned that decisions are made about individuals by AI;
- 40% of survey respondents were concerned about the use of facial recognition; and
- 39% of survey respondents were concerned about targeted online advertising.

Unfortunately, the doteveryone report did *not* engage with society’s understanding or view of the ways in which AI can discriminate. However, in our work, we have spoken to a wide range of people from lawyers to equality bodies to policy makers about AI and how it can discriminate. A common theme to our discussions has been *widespread unawareness* within society of the ways in which AI can discriminate and the extent of the problem. Indeed, the Equinet Report identified that *only* 60% of equality body respondents were aware of a public debate within their country concerning the potential of AI to discriminate.¹⁰ Information from the UK suggests that it is possible that this is changing,¹¹ however given our research across Europe we are doubtful that (if correct) this applies across Europe.

⁸ First White Paper, page 1.

⁹ <https://www.doteveryone.org.uk/report/peoplepowertech2020/>

¹⁰ https://equineteurope.org/wp-content/uploads/2020/06/ai_report_digital.pdf.

¹¹ The CDEI is expected to publish an AI Barometer assessing the attitude of various experts and interested bodies to AI. The authors understand that it has found that there is a widespread understanding that AI can lead to bias and discrimination. More information about the AI Barometer

We stress the lack of awareness within society as to the ways in which AI can discriminate because we suspect that the current widespread mistrust of AI is a view which has been formed with only an *incomplete understanding* of the whole range of risks which AI poses especially to the principle of non-discrimination. We expect that when society better understands the many ways in which AI can discriminate against individuals and is discriminating, then there will be far higher levels of mistrust. On the AI Law Hub we have outlined many examples of the ways in which AI can discriminate.¹²

2. We strongly welcome the Europe Commission's proposal to regulate AI, with a particular focus on equality, so as to create "*trust*" from the public and businesses alike.

Now is absolutely the right time to ensure that appropriate regulation exists within Europe to tackle discriminatory AI so as to ensure that the "*trust*" deficit can be tackled early on before concerns become even more prominent leading to the rejection of the many beneficial ways in which AI can transform business and people's lives. Improved regulation will better enshrine the principle of non-discrimination in the newly emerging field of AI and thereby ensure that the public better embraces it. Appropriate regulation will also provide confidence to business to invest in and deploy new forms of technology. Certainly, our experience has been that businesses with bases in the UK and Europe more broadly perceive the uncertainty around how AI will be regulated within Europe as a significant deterrent to investment and more innovative business models. Moreover, in the absence of clear legal rules, businesses may fear that their competitors will undermine them by deploying AI in unscrupulous ways in the knowledge that there is no meaningful regulatory regime. This approach would again undermine the growth of the economy.

3. Since AI can discriminate in multiple ways, the first step must be to ensure that equality law within Europe covers all sectors (employment, goods, facilities and services) and each of the protected characteristics (sex, age, disability, race, sexual orientation etc).

is available here: <https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation/about>

¹² See <https://ai-lawhub.com/ai-and-discrimination/>

Specific European laws do not currently prohibit discrimination in all areas and in relation to all protected characteristics.¹³ This is a serious omission because AI can and does discriminate in a wide range of diverse scenarios as we have explained on the AI Law Hub.¹⁴ In order to ensure that AI, machine learning algorithms and automated decision making are regulated appropriately, this will need to change. Since AI can discriminate in multiple ways, there must be a right to be free from discrimination in relation to all of the protected characteristics (sex, age, disability, race, sexual orientation etc) when it comes to occupation, the distribution of social advantages and the provision of all goods, facilities and services. In the Equinet Report we have highlighted many examples of good practice taken by states to address this deficit.¹⁵

4. To support a universal principle of non-discrimination, there is merit in introducing targeted procedurally based safeguards with the intention that these will encourage the development and use of systems which comply with the principle of non-discrimination.

A significant obstacle to ensuring the principle of non-discrimination is enshrined within AI systems is the well documented “*black box*” problem. Not only can it be difficult to understand how sophisticated, machine learning algorithms are working, sometimes people do not even know that decisions are being made by about them by a machine.¹⁶

One method of regulation is to introduce procedurally based safeguards. The purpose of these safeguards is to compel organisations to take decisions in accordance with a protocol that will improve the *quality* of the decisions and actions ultimately adopted. In Europe, one example of this type of initiative is Article 7 of Council Directive 2001/23/EC of 12 March 2001 on the approximation of the laws of the Member States relating to the safeguarding of employees' rights in the event of transfers of

¹³ We recognise of course that the principle of equal treatment is a general principle of European Law, which finds expression in the Charter of Fundamental Rights and many other specific pieces of legislation.

¹⁴ See <https://ai-lawhub.com/ai-and-discrimination/>

¹⁵ Op. cit, Chapter 3, Country specific prevention of discriminatory AI and data protection laws, at p. 55.

¹⁶ A powerful example of a system which wholly lacked transparency in this way is the SyRI decision which was recently the subject of litigation in Amsterdam. Our analysis of the case is available here: <https://ai-lawhub.com/2020/03/30/syri-think-twice-before-risk-profiling/>

undertakings, businesses or parts of undertakings or businesses which creates information and consultation rights for employees connected to the transfer of undertakings (**the Acquired Rights Directive**).¹⁷ The Acquired Rights Directive does not require organisations to make certain business decisions which benefit or protect employees, instead it requires employers to inform employees (ordinarily through their representatives) of various matters and consult with them in advance of any transfer. Adopting compulsory procedurally based steps in this way increases the likelihood of decisions considering the voice and perspective of relevant employees and therefore are likely to be less harmful than they might otherwise be for those employees. In short, rather than dictating the outcome, the process around decision-making is carefully controlled.

Similarly, we consider that the introduction of regulation through procedurally based safeguards in the AI sphere would be a means of creating additional transparency, which in turn would create accountability, which in turn would increase the likelihood that the principle of equality was embraced.

5. Procedurally based rules which should be considered as a means of encouraging businesses to comply with the principle of equality are as follows: an AI register (for example, those with “high risk” applications of AI), mandatory AI auditing, a requirement to publish audit documentation and specialised procedural rules concerning the processing of biometric data.

We consider that the following procedurally based rules would promote transparency and therefore accountability and therefore a greater incentive to embrace the principle of equality on the part of organisations businesses and public bodies:

- A register of AI systems which explains which AI systems are deployed by particular organisations along with an explanation of how the system is used and what data it processes. Mechanisms of appeal and challenge should also be clearly signposted. This proposal is very much consistent with the various proposals in the First White Paper around transparency, for example that “... citizens should be clearly informed when they are interacting with an AI system and not a human bring” (page 20). However, we recognise also that choices will be necessary as to the kinds of AI systems that must be registered. This is an issue both of definition and proportionality. It must be clear what should be registered and in making that decision it should be accepted that AI systems

¹⁷ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32001L0023&from=EN>

with minor and wholly beneficial effects might be exempted. Framing appropriate legislation to secure this end may not be easy but it is by no means impossible. We discuss this point further below.

- Mandatory auditing of AI for equality issues pre-implementation and on an on-going basis. This type of process is described as a “*prior conformity assessment*” under the heading “*F. Compliance and enforcement*” in the First White Paper (page 23) although we note that there is no specific reference to testing AI systems for equality implications which should be part of any pre-implementation testing process.
- A requirement to publish audit documentation so that the public can understand the collective impact of AI systems especially on protected characteristics.
- Specialised procedural rules concerning the processing of biometric data such as facial or voice, recognition (or mapping) technologies, so as to ensure the ability to appropriately measure and assess the impact on protected groups.

Specifically, one difficulty which has been encountered in the UK is that data protection rules have led to organisations quickly disposing of biometric data after it has been processed with the unintended consequence that equality monitoring is impeded. A recent example of this problem is the decision of the High Court in the UK in *R v The Chief Constable of South Wales & others ex parte Bridges* [2019] EHWc 2341¹⁸ in which Liberty (a civil liberties group) challenged facial recognition technology deployed in Cardiff. The judgment reveals that South Wales police immediately deleted biometric data from its system once the algorithm had concluded that an individual was not a “*match*” for its watch lists. Whilst this policy is laudable in many ways, it also creates a potential problem: how can it be assessed whether the system is behaving in a non-discriminatory way if data evidencing its decision-making process is not retained? We consider that the answer to this problem is the creation of special rules in relation to biometric data which allows data to be retained and assessed for discrimination at regular, random intervals. The ability to “*dip test*” data at random intervals for equality problems would be consistent with the principle of data minimisation which is central to the GDPR. At the same time, it would allow equality to be effectively monitored.

¹⁸ A copy of the judgment can be accessed here:
<https://www.bailii.org/ew/cases/EWHC/Admin/2019/2341.html>

In short, the need for specialist treatment of biometric data is envisaged by the First White Paper under the headings “*f) Specific requirements for remote biometric identification*” (pages 21 - 22) but we feel that the analysis there does not go far enough to ensure that the principle of equality is respected in this highly sensitive areas.

6. There is merit in limiting new procedurally based safeguards to “*high risk*” applications of AI only

As we have noted, any new regulatory steps in this field of activity must be proportionate. This is not to say that a purely risk-based approach should be taken but rather that any new regulation must take into account the importance of encouraging the development of entirely beneficial technologies. The First White Paper makes this point when it says that regulatory intervention should be proportionate and then proposes a series of measures which will be targeted, it appears exclusively, to AI systems which are determined to be “*high risk*” so as ensure that overly onerous regulation is avoided:¹⁹

... The mandatory requirements contained in the new regulatory framework on AI (see section D below) would in principle apply only to those applications identified as high risk ...

We are conscious that there are many algorithms and AI systems in use that are fairly innocuous. For example, a piece of code which predicts, based on prior usage, which document a worker is most likely to need once a program is opened or which route will most quickly ensure that they arrive at work on time. So as to ensure that these forms of “*low risk*” AI are not subject to onerous rules, we agree that adopting a threshold test of “*high risk*” for mandatory procedurally based safeguarding, rather than universal rules, which ensure that businesses and public authorities were not subject to excessive and overly onerous requirements which might stifle investment and innovation.

7. AI should be classed as high risk where it “*produces legal effects*” for individuals or “*similarly significant effects*” so as to dovetail with Article 22 of the GDPR.

¹⁹ First White Paper, page 17.

The First White Paper proposes twin criteria for determining whether an AI system is “high risk”.²⁰ We propose a slightly modified approach which is use a definition of “high risk” which mirrors the language in Article 22 of the GDPR which prohibits decisions being taken about individuals based on solely automated processing, including profiling, where it “produces legal effects” for individuals or “similarly significant effects” unless certain very stringent exceptions apply.²¹ In other words, Article 22 of the GDPR prohibits AI systems from making decisions about individuals, without any human intervention, where those decisions lead to significant consequences, including but not limited to, discrimination.²² Creating a regulatory regime which imposed greater procedural obligations so as to ensure transparency in relation to AI systems which also fell into Article 22 would have the advantage of creating a cohesive legal system which reinforced the transparency obligations which are contained within the GDPR in relation to data processing (e.g. Articles 5, 12, 13 and 14).

We add a note of caution though. We think it very likely that without careful thought the degree of “human” intervention will become nominal. It is predictable that busy (or bored) humans will simply adopt AI outputs without conscientiously considering and understanding how the system has reached any suggested course of action or its appropriateness in any particular context. Smart regulation will be needed to avoid this occurring.

8. The European Commission should create or support a programme of inquiry so as to ensure that the relevant “high risk” applications of AI are identified.

Our research for Equinet leads us to conclude that adopting a definition of “high risk” which mirrored Article 22 would identify AI systems in a range of fields from recruitment to identification for important services (e.g. passport control) to prediction of risk by or for the state (e.g. risk of committing a crime, risk of fraud, risk of abuse) would be identified as high risk.²³ The European Commission should create or support a programme of inquiry at a national and / or pan European level so as to

²⁰ First White Paper, page 17.

²¹ <https://ai-lawhub.com/data-protection-legal-framework/>

²² See page 21 in the paper “Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679” produced by the European Data Protection Board which is available here: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

²³ See Annex 1 at https://equineteurope.org/wp-content/uploads/2020/06/ai_report_digital.pdf

ensure that all high risk applications of AI are comprehensively identified and then regulated.

9. The introduction of targeted procedural rules is very different to the use of *standards*. The principle of non-discrimination is universal, and the substantive requirements of equality should never be targeted at particular products or sectors using a risk-based approach.

Targeted regulation should never lead to the dilution of the *universal* principle of non-discrimination. We highlight this point because at present, the language deployed in the First White Paper suggests that the European Commission is considering targeted *standards* which will impact on equality, rather than simply targeted procedurally based *safeguards*. For example, it is proposed that there could be “*requirements ensuring that AI systems are trained on data sets that are sufficiently broad and cover all relevant scenarios needed to avoid dangerous situations*” (page 19), “*particular obligations to use data sets that are sufficiently representative, especially to ensure that all relevant dimensions of gender, ethnicity and other possible grounds of prohibited discrimination are appropriately reflected in the those data sets*” (page 19) and under “*d) Robustness and accuracy*” a requirement for AI systems to be demonstrably accurate (pages 20 – 21).

In our view, in order to ensure non-discrimination in AI system, sufficiently diverse data sets and accuracy are an *essential* feature of non-discrimination and should never be limited to certain sectors or products or indeed organisations of certain sizes. Accordingly, we suggest that the European Commission should only entertain targeted regulation in so far as they are procedurally based rules that help to create accountability and transparency in high risk areas. Any standard setting measures, such as a requirement that data sets used for training are sufficiently broad, should be universal in their application.

We acknowledge that there is an argument that ill-considered and over burdensome regulatory rules can advantage business that are well financed and well established when competing with new entrants into their market. This issue is well understood in competition law which has well defined rules to prevent market abuse and for instance to ensure that IP is available on FRAND terms. Nothing we advocate should be understood as seeking to undermine these very important controls on market abuse.

10. A certification scheme should be introduced for “*high risk*” applications which indicate when AI has features that are consistent with the principle of non-discrimination, for example, it utilises a balanced data set or it is “*human centric*”.

We agree with the proposal contained in the First White Paper that a system of certification should be introduced (pages 23 & 24). We suggest that a certification scheme could be used to indicate to consumers that an AI system conforms, in certain ways, with the principle of non-discrimination. Specifically, it is possible that AI systems are too diverse and nuanced to be subject to a certification scheme which certifies that the entirety of the system is free from discrimination.²⁴ However, we foresee the development of a certification scheme which could certify certain aspects of the AI scheme which are consistent with or suggestive of compliance with the principle of non-discrimination. For example, different marks could be used to indicate that the AI system has been trained on a balanced data set or that a human is part of the decision making. Further, limiting such a system to “*high risk*” applications of AI would be consistent with the principle of proportionality.

11. There should be a prohibition on decisions being taken by solely an AI system in a way that mirrors Article 22 of the GDPR.

We fully support the proposal in the First White Paper that regulations should be introduced so as to ensure minimum levels of human involvement dependent on the nature of the AI system (page 21). We suggest that rules dictating minimum levels of human involvement should mirror Article 22 of the GDPR so as to ensure consistency and cohesion within European law (see Section 7 above). Further, it should be made clear that there is a general principle of European public law that decisions should be taken by humans rather than machines unless there has been a specific legislated permission for this not to happen.

12. All organisations involved in the development of AI should be liable for any discrimination including the company, organisation or public body that ultimately uses it.

The First White Paper recognises that numerous actors are involved in the creation and implementation of AI systems from the developer to an end user (page 22). It

²⁴ Although we note that AI systems have received the CE Mark approval from the British Standards Institute: see <https://www.globenewswire.com/news-release/2020/04/15/2016139/0/en/Behold-ai-awarded-CE-Mark-approval-for-its-AI-based-chest-X-ray-diagnosis-technology.html>.

proposes that “in a future regulatory framework, each obligation should be addressed to the actor(s) who is (are) best placed to address any potential risks. For example, while the developers of AI may be best placed to address risks arising from the development phase, their ability to control risks during the use phase may be more limited” (page 22). At this first blush, this sounds like an appealing and proportionate way to regulate AI. However, we have three concerns about this approach in relation to the elimination of discrimination as follows:

- Dividing liability or responsibility along these lines would introduce artifice which would undermine the principle of equality. That is, a body that applies a discriminatory system has offended the principle of non-discrimination in precisely the same way as the body that created the system. To illustrate that point, if company X created a system of selecting employees during a recruitment process that favoured white candidate and sold it to company Y who then used it to only select white candidates, both parties have treated non-white candidates less favourably and offended the principle of equality. In those circumstances, it would be artificial to hold only one company liable (presumably X).
- Beyond the conceptual difficulties with the proposed approach, there is also a practical problem. Specifically, in our experience, some AI systems are developed in partnership with data being supplied by A and processed by B but used by C perhaps with modifications introduced by D. It follows that we consider it unlikely that it will be straight forward to always identify who would be “best placed to address any potential risks”. A regulatory framework can only effectively challenge discrimination in so far as liability can be readily pinpointed with minimal risk of satellite litigation over which person or body is the appropriate defendant to any claim or allegation.
- A further problem is that by placing regulatory obligations only at certain points in the “life cycle” of an AI system, it may disincentivise organisations from taking proactive steps to avoid discrimination. For example, in the hypothetical scenario outlined in the first bullet point, what incentive would there be for company Y to ensure that its system did not discriminate? Potentially none. This problem is exacerbated if company Y is a well-resourced company, possibly based outside of Europe.²⁵ Would a claimant risk litigation against company Y? Maybe not.

²⁵ This complication is recognised at page 22 in the First White Paper.

Accordingly, we consider that there should be appropriate regulation of all actors involved in the development of an AI system.

13. To reflect the difficulties that some end users will face determining whether the AI system is non-discriminatory, due to the “*black box*” problem, end users should be able to rely on a defence that they took all reasonable steps to ensure that the AI system was non-discriminatory. However, this defence should not be available to those who had manufactured or supplied such systems to end-users since they are in a position to ensure that the AI is non-discriminatory.

We propose that the European Commission adopt a different approach towards regulation which is to ensure that actors in the development of an AI system are obliged to act in accordance with the principle of equality but that the end users should be able to avoid liability on the basis that they took all reasonable steps to ensure that the AI system was non-discriminatory. Accordingly, in the hypothetical scenario outlined in Section 12 above, if company Y has taken all reasonable steps to ensure that the product sold by company X is not discriminatory (e.g. regular auditing) then they would be able to avoid liability.

This device is in Great Britain within the Equality Act 2010 whereby an employer can avoid being vicariously liable for the discriminatory acts of its employees provided that it has taken all reasonable steps to prevent the discrimination arising.²⁶ It encourages employers to take steps to proactively minimise the risk of discriminatory behaviour in the workplace such as equal opportunities training.

The defence is highly beneficial to those who might be affected because it introduces a potential second level of regulation that operates at a level close to the source of any problem. It encourages persons using such systems to demand that they are fit for purpose. Moreover we consider it likely that corporate liability insurance policies would drive businesses to take active steps to take advantage of such a defence.

Naturally, this type of defence should not be available to those who had manufactured or supplied such systems to end-users since they are in a position to ensure that the AI is non-discriminatory.

²⁶ S.109 Equality Act 2010, see <https://www.legislation.gov.uk/ukpga/2010/15/section/109/2010-10-01>.

14. To support a universal principle of non-discrimination, the burden of proof should shift to the Defendant where there is a lack of transparency and some evidence to suggest that discrimination is occurring.

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. The reversal of the burden of proof in such situations has been explicitly stated in Directives since 2000.²⁸ There is no reason why this principle would not extend to AI. We propose that the European Commission enshrine within regulation the principle that lack of meaningful transparency as to the way in which an AI systems works, should shift the burden of proof to the organisation deploying the system to demonstrate that discrimination is not occurring. Inevitably, if the service provider is not able or willing to provide such an explanation, then it will not be able to discharge the burden of proof for showing that there is a non-discriminatory explanation for the Claimant's treatment.

At this point we should mention the Second White Paper. This report primarily focuses on how the General Product Safety Directive 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 we can see some useful parallels, there is a degree of artifice in conceptualising AI as simply another type of regulated "*product*" when seeking to ensure that the principle of equality is respected. Despite this limitation, the proposal that there could be a reversal of the burden of proof in relation to harms caused by AI systems (page 14) echoes our proposal here within Section 14. In light of the "*black box*" problem, we consider that there should be a reversal of the burden of proof for AI systems in relation to discrimination claims where there is a complete lack of transparency.

15. The European Union should take steps to ensure that the enforcement of the right to equal treatment (the principle of non-discrimination) and the protection of data rights are fully aligned.

²⁷ C- 109/88, *Handels- og Kontorfunktionaerernes Forbund i Danmark v Dansk Arbejdsgiverforening Ex p. Danfoss A/S* ECLI:EU:C:1989:383.

²⁸ See for instance Article 10 of Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation.

One major issue which we have encountered in our work is the fact that the GDPR was not written expressly to adopt the principle of equal treatment. Data protection regulators therefore tend to think in terms of “*fair*” treatment of data in accordance with the data protection principles and do not start from the premise that no breach of the equal treatment principle is permissible. This has led to some confusion. In our report for Equinet we have emphasised the importance of regulatory alignment and we consider that the absence of this could be a really important obstacle. We have proposed in that paper, some ways in which the GDPR could be brought into alignment with any regulation of AI to prevent discrimination.

In the first place this concerns the extent to which full transparency is required by the GDPR. We have noted the limited extent of the advice from the European Data Protection Board in this respect in the Equinet Report.²⁹ We explain there why we consider this is inadequate to ensure that individuals can assert their right to non-discrimination.

There is also another point we wish to make concerning Article 80 GDPR. This grants data subjects the right to mandate not-for-profits to make complaints about the use of a person’s data. However the provisions are focussed on individual abuse and may not therefore be adequate to assist in bringing claims for breach of the principle of equal treatment where it is often necessary to consider the general impact of an “apparently neutral provision, criterion or practice.”³⁰ In our view provision should be made to enable not-for-profits to take similar action in relation to equality law issues arising from AI. This should start with a similar approach to that taken in Article 80 but go further. It should allow mandated organisations to seek collective data sets on behalf of groups of individuals so as to ensure that data processing is consistent with the principle of equal treatment.

16. The European Union must not permit international trade rules to be developed that in any way undermine the right to equality by immunising intellectual property rights from disclosure when necessary and appropriate for the enforcement of those equality rights.

We are concerned about the secrecy of intellectual property rights in AI systems. We are sure that this concern about preserving trade and commercial secrets is likely to become a major obstacle to ensuring that AI systems are developed which can be examined for their consistency with the principle of equal treatment. This point is raised in the First White Paper at section H. We consider it critical that the relevant

²⁹ Op. cit. Chapter 3 ‘Using the GDPR to open ‘the Black Box’ ’, p.53.

³⁰ See for instance the standard definition of indirect discrimination in Article 2 of the Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation.

Directorates of the European Commission co-operate fully to ensure that this does not happen by default and that it is always clear that the international trade rules that apply to intellectual property rights do not undermine equality.

Robin Allen QC & Dee Masters

www.cloisters.com

www.ai-lawhub.com

10th June 2020