EMPLOYMENT
LAWYERS
ASSOCIATION

PO Box 1609
High Wycombe
HP11 9NG

TELEPHONE 01895 256972
E-MAIL ela@elaweb.org.uk
WEBSITE www.elaweb.org.uk

ICO Consultation on Guidance on the AI auditing framework

Response from the Employment Lawyers Association

1 May 2020

ICO Consultation on Guidance on the AI auditing framework

Response from the Employment Lawyers Association

1 May 2020

Introduction

The Employment Lawyers Association ("ELA") is a non-political group of specialists in the field of employment law and includes those who represent claimants and respondents in courts and employment tribunals. It is not ELA's role to comment on the political or policy merits or otherwise of proposed legislation or regulation, rather it is to make observations from a legal standpoint. ELA's Legislative and Policy Committee consists of experienced solicitors and barristers who meet regularly for a number of purposes including to consider and respond to proposed legislation and regulations.

A working party, co-chaired by David Widdowson and James Davies was set up by the Legislative and Policy Committee of ELA to respond to the ICO consultation on Guidance on the AI framework. Members of the ELA working party are listed at the end of this paper.

Data protection laws impact significantly on many areas of employment law and, as a consequence, ELA members will generally need to have an understanding of data protection principles.

Artificial intelligence (AI) is rapidly increasing in employment-related decision making. For the time being AI is concentrated in recruitment decision-making but it is inevitable that its use will become more common in other areas of employment-related decision-making before long.

In this survey response we focus on those parts of the consultation paper most relevant to employment lawyers.

Q1 Is the draft guidance clear about what you should consider when creating and using AI-systems that are compliant with data protection law?

- ☐ Yes
- ☐ No

Please outline what parts, if any, you think could be improved:

Q2 How well-pitched are the sections in the draft guidance?

a 'About this guidance'

- ☐ Too detailed
- ☐ Just right
- ☐ Too vague

Please provide your suggestions on how we can improve on the level of detail:

b 'What are the accountability and governance implications of AI?

- ☐ Too detailed
- ☐ Just right
- ☐ Too vague

Please provide your suggestions on how we can improve on the level of detail:

In the employment context, we can see a real dilemma in determining the required accountability of employers for algorithms used to make employment-related decisions. We support the ICO in addressing this issue. We do, however, consider that this requires a joined-up approach both internationally and with other domestic regulators (see below).

Employers may develop their own systems or may work with a developer in producing a bespoke algorithm. However, in many cases, the employer will procure a third-party system often developed outside of the jurisdiction (and very often from the US). Employers will also look to use these systems across jurisdictions.

We explore in more detail below the particular challenges for employers in ensuring that AI-based employment decisions do not discriminate unlawfully under UK equality laws. Currently, the legal framework presents significant evidential difficulties for individuals who seek redress where they consider they may have been discriminated against as well as to employers who seek to show that they have complied with their legal obligations.

We consider it vital the guidance set out clearly the steps an employer is expected to take which are realistic such that UK employers are not, in practice, prevented from embracing the irresistible advance of AI whilst properly complying with their obligations not to unlawfully discriminate and protecting individual rights.

It is usually not possible to expect the employer who has acquired an off the shelf product to have the resources or expertise to satisfy itself that the algorithm does not discriminate. This fact reveals a tension in the framework. Controllers are responsible for their own compliance, even if they choose to adopt a vendor's product wholesale. Lack of resourcing would be unlikely to excuse an organisation from conducting a data protection impact assessment/consulting with the ICO etc if required. Such an assessment could involve assessing potentially discriminatory impact of processing by an off the shelf AI product, around for example the lawful use principle under the UK GDPR. This is because an AI product whose processing produced unlawfully discriminatory outcomes would be unlikely to be grounded in any of the lawful bases for processing.

The guidance should explain that, even with extensive training and thorough diligence procedures in place, it is likely be too complicated for businesses to understand how candidate personal data is being collected and processed by third party products and how decisions are being applied to that data even if they had access to the detailed code and datasets which is again unlikely as it will normally be a trade secret.

We are concerned that the current expectations set out the guidance are not sufficiently realistic so that compliance is likely to mean the end of most of the use of AI in employment-decisions. We do, nonetheless, believe that realistic expectations can be set for employers which ensure individual rights are protected. The ICO cannot, of course amend, controllers' statutory obligations, but may publish regulatory enforcement policies and has allowed regulatory 'sandboxes' in controlled settings. Discrimination law and the UK GDPR are not co-extensive regimes but intersect around matters such as the lawful use principle, as set out above.

It can therefore be a challenge for businesses to have real certainty about the extent to which there is any underlying discrimination in the results produced. The guidance should underline this challenge and provide practical advice on how to manage the risk of procuring third party products against their potential benefits and value to the business.

By way of illustration, on page 50 of the draft guidance, the following statement is problematic: 'Ensure that all functions and individuals responsible for its development, testing, validation, deployment and monitoring are adequately trained to understand the associated statistical accuracy requirements and measures'. It would be better to say that associated statistical accuracy requirements and measures must be brought to bear on the situation. If this is required, it is very likely by way of consultation with a suitable expert.

The draft guidance does say that businesses should not "*delegate to data scientists or engineering teams*" (page 13) and underlines the accountability of senior management on several occasions. However, it should also stress the need to train those who are using and instructing providers of AI systems day to day, so that they can flag issues and concerns as and when they arise. Whilst we would accept that decisions as to how AI systems are applied must rest with operational management equally those decisions should not be taken without recourse to appropriate specialist advice.

It may be that some form of kitemark system might be devised so that proprietary systems and/or those developed in-house by employers can demonstrate compliance with equality laws.

From a recruitment perspective, this means training HR and recruitment teams, as well in-house lawyers and compliance teams (who can advise during the procurement stage of a third-party AI product for example). Controllers may of course be required to appoint a data protection officer.

c 'What do we need to do to ensure lawfulness, fairness, and transparency in AI systems?'

- ☐ Too detailed
- ☐ Just right
- ☒ Too vague

Please provide your suggestions on how we can improve on the level of detail:

Employment law and data protection principles converge most obviously in the area of "lawfulness, fairness and transparency". In particular, employment lawyers are already grappling with the application of discrimination law (Equality Act 2010) to AI-based decisions. Under the Data Protection Act 2018, there are further considerations for dealing with special categories of data for trade union membership which cross-over with separate employment law rights against unlawful inducements, detriment and dismissal under the Trade Union Labour Relations (Consolidation) Act 1992 that are underpinned by international human rights law.

AI-based decisions also bring into play contract law (for example, the implied term of trust and confidence) and other statutory rights which affect the workplace (for example, health and safety legislation). However, the for the time being, the issues are greatest with discrimination law not least as AI is, to date, used most frequently in recruitment decisions where individuals' employment law rights are largely restricted to discrimination law.

Discrimination and AI

ELA welcomes the ICO's attention to this area as it is fraught with problems. We are, as mentioned above, concerned that some controls set out on pages 62 and 63 of the draft guidance are unrealistic for many employers using AI for recruitment decisions.

Risks of discrimination

The risk of discrimination arises at all stages of the development and use of AI bias can and does arise in the entire machine learning pipeline. It includes bias arising from the demographic of practitioners and developers, and crucially where human judgement is used to label training or validation data.

The training set can be tainted; the defined outcome can result in discrimination; the predictors developed by the algorithm can result in unlawful discrimination; the validation set can be tainted; and the data inputted in respect of each candidate can result in discrimination.

The draft guidance highlights the potential for discrimination if an AI system is trained on a dataset containing underlying and/or subconscious bias or past discrimination. In the recruitment context, the draft guidance includes the example of an AI system used to score job applications, trained on examples of previously successful candidates (page 54). The draft guidance explains that the system may learn to reproduce any historic discrimination in selecting candidates.

While consideration of the training data (and, equally, the validation data) put into an AI system is crucial in managing the risk of discrimination, the draft guidance should also underline the importance of carefully defining the parameters of the required output. In the recruitment setting, such parameters would be the criteria for a successful candidate.

If an employer defines a role too narrowly, for example with performance criteria that may favour men over women, the results of a search using an AI system may be skewed narrowly too as the system tries to find the closest match to the employer's ideal candidate. Continuing this example, an AI system could therefore find that men are a closer match than women to a particular role – thus perpetuating an underlying bias (even if the training data has been carefully considered).

To mitigate this risk, the Guidance should underline the importance of quality input data and due consideration to the specified output. In a recruitment context, this could mean: defining job descriptions as openly and as neutrally as possible; and aligning role criteria with a business's forward-looking diversity strategy, rather than based on traits of previous successful candidates.

The guidance should also emphasise the risk that the predictors are flawed or discriminatory. A predictor might be accurate but nonetheless discriminatory.

Employer's obligations

We agree with the draft guidance which recognises that it must be the employer who will normally be the data controller to ensure that the lawfulness, fairness and transparency of the processing of personal data. Nonetheless, the key question is the extent of the employer's obligations in meeting this threshold. It raises a number of supplemental concerns:

Can it ever be enough to rely on the developer's warranty to satisfy this obligation and, if so, what does the employer need to do to satisfy this?

Can it ever be enough to rely on independent verification that the algorithm does not discriminate unlawfully?

Should it always be the case, that the employer must satisfy itself that the algorithm produces lawful outcomes? Should a smaller employer go to the same lengths as a large employer to satisfy itself that the algorithm produces lawful outcomes?

Should the employer's duty differ if it has developed or commissioned a bespoke product from cases where the employer has acquired an off the shelf product?

Whatever route is taken, demonstrating that an AI-based decision is not unlawfully discriminatory is hugely problematic and any draft guidance needs to acknowledge this.

Verification/mitigating risks

Many developers of off the shelf algorithms will promote their products as having been stress tested to avoid discrimination. However, the algorithm and the testing will usually be trade secrets that neither the employer nor an aggrieved individual will easily get hold of the test data and results. As mentioned above, these developers are often outside the jurisdiction of the tribunal. This is further complicated where deep-learning "black box" algorithms are used.

The verification can either be "algorithm focused" or "outcome focused". The draft guidance suggests three approaches to mitigating the risks of discrimination in machine-learning models: (1) anti-classification (algorithm focused); (2) outcome/error parity (outcome focused); and (3) equal calibration (outcome focused).

Both are algorithm focused and outcome focused verification are problematic. Algorithm focused verification entails an analysis of the algorithm to ensure potentially discriminatory characteristics are removed. This is less easy for complex algorithms or deep-learning algorithms which are continually evolving.

Outcome focused verification is based on misconception that (at least in the UK) equality of outcome proves an absence of discrimination. Algorithm fairness does not equate to non-discriminatory algorithms. For example, the example given at the bottom of page 53 illustrates the point. The emergence of lower credit scores for women does not prove, of itself, that there is a difficulty with bias or what that difficulty is. The draft guidance confuses statistical inference with equality. The example assumes (incorrectly) that equality of outcome must signal a lack of bias in the data.

This issue is further complicated by the wide range of protected characteristics which are covered by discrimination laws and related legislation that protects special categories of data. Even if these outcome tools were of use an exercise would need to be completed for each characteristic. The suggestion that using separate training models for men and women is a valid means of mitigating discrimination should be removed. As the draft guidance acknowledges (page 55), the creation of different models is likely, in itself, to violate non-discrimination laws in the UK. Further, as mentioned above, models would be needed for each protected characteristic not just gender.

None of these approaches can ever adequately show an absence of unlawful discrimination and related legislation that protects special categories of data. They might mitigate the risk to the extent that they might reduce the risks of discrimination but never eliminate it. Further, particularly with outcome focused models, they might equally create discrimination for a group with a protected characteristic over another group with a different protected characteristic. This should be made very clear in any draft guidance.

The ideal in machine learning terms would be a hygienic training dataset, one where the influences of unlawful discrimination are not present. For reasons that are given in the draft guidance, when using 'real' data, for example members of a given workforce to train as CV checking system, the files of an insurance company to predict claim rates, and so on, this cannot be achieved. The commercial reality is even more complex. Machine learning is an ongoing process whereby the AI tool is frequently running over an increasing/evolving data set to find patterns and correlations. This data is procured on an ongoing basis, and from multiple sources (employers; recruiters; online platforms). So, in reality, there is not a fixed point in time where both the various parts that comprise the training data, and the relevant AI tool, can realistically be said to be free of bias. It is a moving feast.

UK discrimination law

We wonder if the approach set out in the draft guidance is being driven by compliance with US discrimination laws.

In the UK and the EEA any algorithm-based employment decision will have to comply with a wide range of laws such as those on direct discrimination; indirect discrimination; and the duty to make reasonable adjustments.

It seems to us that the term "proxy" is used to cover two different situations and we suggest that this is addressed. In cases of direct discrimination, a person is treated less favourably because of their protected characteristic. As set out in the draft guidance (page 55), even if a person's protected characteristic is not available to the algorithm, it may be possible to determine this from other information (the proxy). The example in the draft guidance is of attendance at a single-sex school. Direct discrimination (other than in cases of age discrimination) is, generally, always unlawful.

In cases of indirect discrimination, an apparently neutral provision, criterion or practice (PCP) (the proxy) is applied to a group and puts persons who share a protected characteristic within that group and the person alleging discrimination, in particular, at a disadvantage.

Unlike direct discrimination, indirect discrimination is not always unlawful. It is capable of being justified where the use of the provision, criterion or practice can be shown to be a proportionate means of achieving a legitimate aim.

As the draft guidance highlights (page 55) removing all possible PCP proxies may leave very few predictively useful features. There is no need to eradicate all such proxies provided they can be objectively justified.

It is unclear how an algorithm might go about assessing whether or not such a proxy can be objectively justified. It is also probably unrealistic to expect the employer using an algorithm to identify the potentially indirectly discriminatory features and assess whether or not they can be objectively justified, particularly with off the shelf products where the employer does not have access to the code.

It is also unclear how an employer might satisfy itself that an algorithm is complying with the legal duties to make reasonable adjustments e.g. under disability discrimination law. These duties mean that an employer must discriminate positively in favour of an employee with a disability.

The guidance does not distinguish between lawful and unlawful discrimination.

Employee data

Verifying an absence of discrimination on any of the protected characteristic requires the inputting of the relevant information in respect of the individuals in any verification set or the affected individuals. Without this data, employers and system providers have no realistic means to: (a) assess whether the outputs of an AI system are potentially discriminatory (or not) against persons with any of the protected characteristics; or (b) decide what measures might be required to offset any discriminatory outcomes.

Indeed, the draft guidance contains an example (page 57) which states that an employer could collect details of a sample of candidates' religious beliefs for these purposes. However, in reality, candidates (and employees) often do not provide data regarding their protected characteristics, even when asked. They are under no legal obligation to do so. They are (understandably) sensitive and sometimes untrusting about providing such sensitive data. This problem has been recognised by the UK Government in its consultation paper regarding ethnicity pay reporting. Unless employers and/or system providers can find a way to collect all of the data needed, we query whether they can take any steps to assess and/or mitigate the potential discriminatory impacts of an AI system.

The ICO may take the position that its guidance is not the appropriate place to address these points in detail. However, we offer our view that it would be remiss not to refer in the guidance to the substantial efforts and planning involved in collating data regarding the protected characteristics of candidates and employees.

Legal claims

To date, we have not seen legal claims based on allegedly discriminatory algorithms. However, there is good reason to predict these will become common:

1. Increased use of algorithms;
2. Use of algorithms for employment decisions outside of recruitment (discrimination claims arising out of recruitment decisions are rare in any event);
3. Unlike with human-based decisions, it is (theoretically, at least) possible to identify the true basis of the decision (AI decisions may be less biased and less discriminatory than human-based ones but it may, nonetheless, be easier to demonstrate that bias);
4. For the time being anyway, "algorithm aversion" where individuals are less trustful of AI than human decision-making;
5. Lack of transparency where the basis for the decision cannot necessarily be explained to the affected individual; and
6. Potential claimants recognising the practical challenges for employers defending claims (e.g. an unwillingness or inability to disclose trade secrets).

It is worth highlighting the potential benefits from AI-based decisions. Used properly AI can be a tool for reducing bias in decision-making and it should always be borne in mind that an AI based decision process may not eliminate bias entirely but may be less biased than human-based decisions.

The challenges for an employment law system ill designed for AI-based decision-making will be profound. In our view, the risk mitigation approach outlined in the draft guidance is somewhat unrealistic and flawed.

Possible way forward – statutory guidance

We consider that it is essential that the UK's regulators work together to see how guidance can be available in one place in a form which is legally correct and simple to access. In our view the ICO should work alongside any other regulators with functions concerning AI systems, such as the Equality and Human Rights Commission (EHRC). And, perhaps, the quasi– regulator the Centre for Data Ethics and Innovation (CDEI) to create statutory guidance that explains how employment law applies to the new world of AI.

International co-ordination

In addition, to the extent that it is not already, we recommend that the ICO consult with the Commission Nationale de l'Informatique et des Libertés in France and other relevant supervisory authorities on the substantial public interest processing in the context of AI.

To illustrate the need for such consultation, we note differing approaches across GDPR-covered jurisdictions to the processing of ethnicity data for the purpose of equal employment opportunity (EEO). UK data laws specifically contemplate EEO measures providing a basis to process ethnicity data, whereas as certain other jurisdictions severely curtail or prohibit such processing.

The guidance should recognise that an AI system may, therefore, engage in entirely lawful processing in the UK in a way which would be prohibited elsewhere. Multi-national organisations are left with a choice of applying differential approaches by jurisdiction (which is unlikely to be of value in the AI context) or taking a 'highest common denominator' approach and declining to allow the processing at all across its jurisdictions. The latter may lead to sub-optimal outcomes from a UK perspective, as the EEO measures may have to cease.

We recognise that in this post-Brexit world for some it may seem unnecessary to consider European law; yet this approach has many dangers, not least because so many businesses operating in the UK also operate in some or all of Europe. We are aware from research across Europe, that such businesses will not want to use several different AI systems, i.e. one developed for use in the UK and another for the rest of Europe. They will have a common approach towards AI systems rather than a UK-centric approached. The need for alignment is made all the more pressing by the push for an expedited adequacy determination by the European Commission as it assesses the UK's data protection framework. Notably two British jurisdictions with adequacy status (Jersey and Guernsey) take a more protective stance than even the GDPR, defining criminal data as a species of special category data. This suggests a direction of travel for the UK in its goal to achieve adequacy.

Goods, facilities and service

Employers and the providers of goods, facilities and services are all subject to the Equality Act 2010 in broadly the same way. Accordingly, statutory guidance should bring together, in a coherent way, the employment field and this broader context.

Liabilities

Further, beyond practical guidance, clarity is also required in relation to basic legal concepts and this could be incorporated within a statutory guide.

For example, the extent to which employers can rely on the "statutory defence" within the Equality Act 2010 if they purchase discriminatory technology is currently clear. An employer has a defence to a discrimination claim if it took all reasonable steps to prevent an employee from contravening discrimination law. One helpful way forward would be to provide guidance as the steps which an employer might need to take, say, to ensure independent verification of that an AI-based decision was not unlawfully discriminatory perhaps in the form of a "kitemark" referred to above.

The draft guidance might also highlight the potential liability of the developer under the Equality Act. A third party can be liable where it causes or induces another to contravene the legislation or helps that third party to contravene the legislation.

We think, therefore, that there is a key need for practical guidance as to how business is required by law to navigate these various legal frameworks when operating AI systems. The best form of such guidance is statutory guidance which is frequently produced in the employment field. In our view a practical document focusing on AI is urgently required which covers all these issues. Anything less than this might cost jobs or restrain the development of new ideas or both.

d 'How should we assess security and data minimisation in AI?'

- ☐ Too detailed
- ☐ Just right
- ☐ Too vague

Please provide your suggestions on how we can improve on the level of detail:

e 'How do we enable individual rights in our AI systems?'

- ☐ Too detailed
- ☐ Just right
- ☐ Too vague

Please provide your suggestions on how we can improve on the level of detail:

See above

3 Is it easy to find information in the draft guidance?

- ☐ Yes
- ☐ No

Please provide your suggestions, if any, on how the structure could be improved:

4 Are the risk statements and the examples of controls useful?

☐ Yes

☐ No

Please provide any suggestions, if any, on how these could be improved:

See above

5 Do you have any examples of using the draft guidance in practice that you think would be useful for us to know?

☐ Yes

☐ No

If yes, please provide further details:

6 What industry is your organisation in?

Professional body

7 Do you develop AI in house, or provide/procure it to/from others?

Multiple options allowed

☐ We procure AI from a third party

☐ We create and use AI in-house

☐ We provide AI to a third party/parties

☒ N/A

If yes, please provide further details:

ELA Working Party

Robin	Allen QC	Cloisters	
Kalila	Bolton	Allen & Overy LLP	
James	Davies	Lewis Silkin	Co-chair
Jonathan	Exten-Wright	DLA Piper UK LLP	
Philip	Lindan	Withers LLP	
Dee	Masters	Cloisters	
David	Palmer	Fieldfisher LLP	
Bruce	Robin	UNISON Legal Services	
Michael	Whitbread	HSBC Bank Plc	
David	Widdowson	Abbiss Cadres	Co-chair