
Article:
Atkinson, J. orcid.org/0000-0001-5207-2231 (2020) Automated management, digital discrimination, and the Equality Act 2010. Green's Employment Law Bulletin (159). pp. 3-6. ISSN 1352-2159

Automated management, digital discrimination, and the Equality Act 2010

Dr Joe Atkinson

University of Sheffield, School of Law

Introduction

The recent A-level results fiasco thrust the issue of algorithmic decision-making, and its potential to create unfairness and intensify inequalities, into the public consciousness. Decisions relating to important aspects of our lives are increasingly made by computational algorithms; determining the university we attend, the jobs we get, our access to financial and other services, and our treatment by the state. At the same time, however, empirical research has consistently demonstrated that automated decision-making can lead to injustice and discrimination (for an introduction see O'Neil, *Weapons of Math Destruction* (2016)).

One significant dimension of this is the growing use of algorithmic decision-making by employers to automate managerial functions. Whether existing employment law can protect against unfairness and injustices caused by this 'automated management', and if new regulatory frameworks are needed, are critical questions for employment lawyers.

This article contributes to this broader agenda by considering employers' liability for discriminatory automated management under the Equality Act 2010 (EqA). The analysis is limited to direct and indirect discrimination for sake of brevity. Although there is yet to be case law directly addressing these issues, it is argued that existing anti-discrimination protections are, to a large extent, capable of capturing digital discrimination.

Algorithms, automated management and digital discrimination

An algorithm is simply a series of steps or processes applied to achieve a goal, and 'algorithmic decision-making' generally refers to the application of algorithms by a computational decision-making system. These often now use advanced 'machine learning' techniques which operate with minimal human supervision or instruction, and develop predictive models using patterns identified in existing 'training data'.

Employers are increasingly using algorithms to automate managerial decision-making processes, a practice described here as 'automated management'. At present, the most common use of automated management is in recruitment, for instance to identify strong applicants or analyse video interviews. However, algorithms are also being used to automate other managerial functions, including evaluating performance, setting remuneration, selecting employees for promotion, and triggering dismissal or disciplinary procedures. Many of these techniques were pioneered in the 'gig economy' but are now being adopted more widely.

Employers regard automated management technologies as a means of increasing the speed and quality of decision-making, but algorithmic decisions can also be discriminatory.

There are several potential causes of this 'digital discrimination'. It may result from the design process if the biased assumptions or choices of the algorithms' developers become reflected in the model. At its most direct, an algorithm could rely on personal characteristics such as race, religion or gender as part of its decision-making process. But even where such characteristics do not feature directly, algorithmic decisions may be based on combinations of other factors that amount to close proxies - such as post code and educational history acting as a proxy for race.

More subtly, algorithms may discriminate if they are developed using data that contains historical bias or discrimination. Such algorithms are likely to reproduce, and potentially amplify, inequalities present in the training data. For example, an algorithm trained to identify candidates for promotion using data about a companies' senior management team, which is overwhelmingly white and male, is likely to end up favouring employees from these groups.

One frequently discussed issue with algorithmic decision-making is the lack of transparency over how decisions are made. The opacity of some algorithmic models means that even those developing them are unable to explain the process by which a decision has been reached. This 'black box' problem may appear to make it difficult to hold employers accountable for their use of automated management because the reasons underlying a decision are hidden from scrutiny. As we shall see, however, this is not necessarily a barrier to liability under the Equality Act 2010.

Liability for direct digital discrimination

Direct discrimination occurs where a person is treated less favourably because of a protected characteristic (EqA s.13). The relevant characteristics are: age; disability; gender reassignment; marriage and civil partnership; pregnancy and maternity; race; religion or belief; sex; and sexual orientation (EqA s.4). Employers are prohibited from directly discriminating throughout the employment relationship (EqA s.39). Except in cases of age discrimination, and some limited statutory exceptions, direct discrimination cannot be justified.

Following this, employers will be liable for direct discrimination if they adopt automated management technologies that rely on data about individuals' protected characteristics. Such situations are closely analogous to *Test-Achats* Case C-236/09, which found it was unlawful discrimination for companies to use sex as part of their calculations when determining the price of insurance.

Cases where protected characteristics feature explicitly in algorithmic decision-making will (hopefully) be rare. But there could also be liability for direct discrimination if an algorithm is programmed to ignore protected characteristics, but in effect fails to do so because it relies on data points that act as proxies with 'exact correspondence' to them (*R (Coll) v SoS for Justice* [2017] UKSC 40).

Employees who are treated less favourably by an algorithmic decision-making tool that relies on a protected characteristic, or a precise proxy, will therefore have a claim for direct discrimination. Crucially for cases of discrimination-by-algorithm, there is no need for the discrimination to be intentional or malicious. As set out in *R (E) v Governing Body of JFS* [2009] UKSC 15, the question is whether a protected characteristic is the reason for the decision, and the guidance in *Igen Ltd v Wong* [2005] IRLR 258 makes clear the treatment must 'in no sense whatsoever' be based on the protected characteristic.

Despite this seemingly strong prohibition, claimants will often struggle to prove automated management technologies are discriminatory because they will not have access to the algorithm's internal reasoning. While significant, this need not be an insurmountable barrier to liability. Much will depend on the operation of the burden of proof.

In recognition of the difficulty claimants face in proving discrimination, EqA section 136 requires that courts find discrimination wherever there are 'facts from which [they] could decide, in the absence of any other explanation,' it has occurred. A two-stage approach to s.136 was confirmed in *Royal Mail Group Ltd v Efobi* [2019] EWCA Civ 18, whereby claimants must initially demonstrate facts from which discrimination can be inferred, and the burden then shifts to the respondent to demonstrate the treatment was on non-discriminatory grounds.

The threshold for claimants to satisfy the initial burden of proof, i.e. the facts from which courts are willing to infer discrimination, will be key in digital discrimination cases. It is not normally enough for claimants to show they have been treated less favourably than another person who does not share the protected characteristic. As discussed in *Madarassy v Nomura International plc* [2007] ICR 867, there must be facts from which the court can infer there *was* discrimination, rather than merely suggesting that there *could* have been. In cases involving automated management, however, the empirical evidence of widespread bias in algorithmic decision-making systems might provide the court with sufficient facts to infer discrimination. Certainly, if the same technology or process used by an employer has already been proven to be discriminatory in another context then this should be sufficient to meet the burden of proof.

Alternately, claimants may be able to prove discrimination by seeking disclosure of information about the algorithm's outputs, training data, or internal reasoning processes. Discrimination can be inferred from a historical pattern of individuals with a protected characteristic being treated less favourably, as in *Rihal v London Borough of Ealing* 2004 IRLR 642, so the burden of proof would likely be met if the algorithm's output data revealed a pattern of this kind. Similarly, if the algorithm's training data is shown to be biased this should be enough to infer discrimination, because the bias will likely be replicated in the resulting model. In addition, refusal by employers to be transparent about the algorithm's output data or reasoning process may also lead to an inference of discrimination (see *Danfoss* Case C-109/88).

If a *prima facie* case of direct digital discrimination is demonstrated, the burden shifts to employers to prove that treatment was not 'because of' a protected characteristic. At this stage the 'black-box' nature of automated management could be problematic for employers, as the

complexity and opacity of algorithmic decision-making may make it difficult to show that automated management technologies do not rely on protected characteristics or close proxies.

Indirect digital discrimination

Indirect discrimination occurs where a seemingly neutral 'provision, criterion or practice' (PCP) is applied, which in fact puts a group sharing a protected characteristic at a 'particular disadvantage'. Employers who apply indirectly discriminatory PCPs at any stage of the employment relationship will be liable to members of the disadvantaged group who suffer the disadvantage, unless the PCP can be justified as a proportionate means of pursuing a legitimate aim (EqA s.19).

Automated management will more commonly lead to indirect discrimination than direct, because it is easier to ensure algorithms ignore protected characteristics than to prevent them disadvantaging protected groups. It is therefore particularly important that the law protects against this form of digital discrimination.

An inclusive approach is taken to defining PCPs, meaning the use of automated decision-making by an employer would undoubtedly be a 'practice' for the purposes of indirect discrimination. Employers could therefore be liable if their use of automated management technologies puts a protected group at a particular disadvantage.

To prove a 'particular disadvantage', claimants must show a disparity of impact between the group sharing a protected characteristic and the general population the PCP is applied to. Automated management may create this disadvantage in a number of ways. For example, if members of the protected group are overrepresented in the class of people detreimentally impacted by automated decision-making, or if they are statistically less likely to benefit from decisions, or are subject to a higher rate of errors. As with direct discrimination, claimants can seek disclosure of the algorithm's decision-making in order to demonstrate a discriminatory impact, and refusal by employers may lead a to an inference of discriminatory impact.

Importantly however, the Supreme Court in *Essop v Home Office* [2017] UKSC 27 found that claimants need not demonstrate *why* a PCP creates a particular disadvantage; allowing claimants to sidestep the potentially problematic 'black box' nature of algorithmic decision-making.

Once an employers' use of automated management is proven to have a discriminatory impact, the burden will shift to them to justify the practice. There are three stages of the justification test: the use of automated management must pursue a legitimate aim; it must be capable of achieving that aim; and it must be (reasonably) necessary and proportionate (for discussion see *Homer v Chief Constable of West Yorkshire* [2012] UKSC 15).

Employers will invariably pass the first stage as they will be able to point to real business reasons for adopting automated management practices. They will also likely pass the second stage, providing they can show the technology is functioning effectively and accurately. However, the requirement of necessity and proportionality may prove more difficult.

At this final stage, courts must balance the discriminatory impact of the PCP against the employers need to achieve their aim; the greater the harm and number of employees affected, the more difficult it will be to justify. The availability of protection against indirectly discriminatory automated management practices will therefore largely turn on tribunals' assessment of necessity and proportionality on the facts of each case, and the strength of scrutiny applied here will be key. Significantly, however, if the employers' aim could be achieved by a less discriminatory means, for example by adapting the algorithm, this may lead to a finding that the PCP goes beyond what is reasonably necessary so cannot be justified.

Conclusion

Employers' use of automated management and algorithmic decision-making is only likely to increase in the coming years, despite the serious risks of workplace discrimination. It is therefore welcome that, although necessarily somewhat speculative, the above analysis indicates the Equality Act 2010 has the potential to protect against these emerging forms of discrimination. However, the ability of anti-discrimination law to capture instances of digital discrimination remains to be seen, and will largely depend on the courts' approach to the issues of burden of proof and justification.