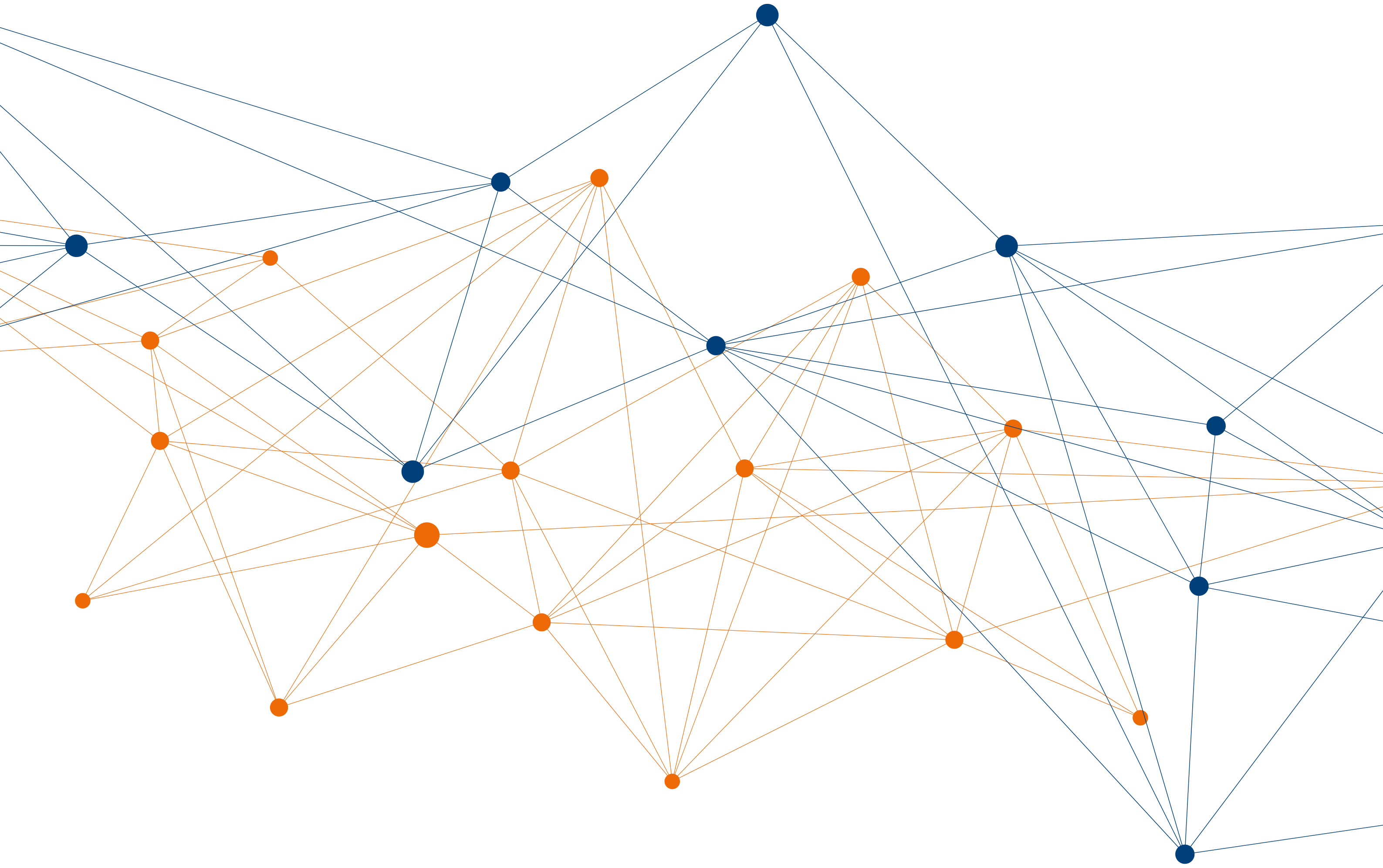
Institute *for the* Future of Work

Artificial intelligence in hiring

Assessing impacts on equality



Putting people first

Authors

Logan Graham
Research Fellow, IFOW

Dr Abigail Gilbert
Principal Researcher, IFOW

Joshua Simons
Research Fellow, IFOW

Anna Thomas
Director, IFOW

Published April 2020

With Helen Mountfield QC

Executive summary

The use of artificial intelligence (AI) presents risks to equality, potentially embedding bias and discrimination. Auditing tools are often promised as a solution. However our research, which examines tools for auditing AI used in recruitment, finds these tools are often inadequate in ensuring compliance with UK Equality Law, good governance and best practice.

We argue in this report that a more comprehensive approach than technical auditing is needed to safeguard equality in the use of AI for hiring, which shapes access to work. Here, we present first steps which could be taken to achieve this.

This work has been completed as part of the Institute for the Future of Work Equality Programme.

Key findings

- Auditing tools are rarely explicit about their purposes – users need to understand what they are evaluating, and why
- Auditing tools are rarely explicit about key definitions of bias or fairness
- Auditing tools made in the US routinely import US assumptions about the requirements of Equality law, which differ in the US and the UK
- Auditing tools offer a ‘snapshot’ of bias in an AI system, rather than an evaluation of its impacts over time
- Impacts of AI systems on equality are not adequately considered, or prioritised, within existing approaches to auditing
- Auditing tools are not designed or equipped to address and mitigate many forms of bias, discrimination and inequality when they are detected
- Unless auditing tools are focussed on relevant equality questions, and their use is integrated into a wider equality impact assessment, their utility for promoting equality is limited

We recommend

To companies:

- Equality should be recognised as a guiding principle in the deployment of AI and auditing systems, alongside Fairness, Accountability, Sustainability and Transparency and Data Protection principles
- Regular auditing for equality, and taking steps to make appropriate adjustments where inequality is identified, are required to avoid breach of the Equality Act where AI systems are used to determine access to work
- Companies should integrate technical auditing into a wider equality impact assessment help understand the different types of impact on equality and take action in response
- To promote legal compliance, good governance and best practice, this wider equality impact assessment should aim to exceed the requirements of national equality legislation, as well as data protection and employment legislation
- Before deploying automated hiring tools, companies should consult their workforce and any affiliated union, to discuss potential impacts on equality and proposal for equality impact assessment

To policy-makers:

- Equality should be recognised as a guiding principle in the design and deployment of AI and auditing systems, alongside Fairness, Accountability, Sustainability and Transparency and Data Protection principles
- CEOs and HR leaders need practical guidance on effective auditing and its wider role to promote legal compliance, good governance and best practice
- Professional and industry standards for auditing tools, including auditing for equality, are urgently required to maintain high, consistent standards. We recommend that this initiative is led and coordinated by the CDEI
- Auditing must fit within a broader approach to evaluating the impact of AI systems on equality. This comprehensive evaluation should include reasonable consideration of impacts on equality of opportunity and outcome, and focus companies on the making of adjustments to mitigate relevant adverse impacts which have been identified
- Equality impact assessments would provide insight to inform collective debate about possible proactive steps by employers and others to actively promote equality between individuals and groups

To developers:

- An improved technical auditing tool should be developed which pays close attention to the requirements and purpose of UK Equality Law
- Computer scientists and policy makers should work together to develop (i) a greater understanding of the risks presented by AI to patterns of systemic inequality over time (ii) develop approaches to addressing bias, fairness and equality
- Equality should be recognised as a guiding principle in engineering and design, alongside Fairness, Accountability, Sustainability and Transparency and Data Protection principles
- All auditing tools should:
  - Define and share sensitive attributes used for auditing;
  - Define and share in words all key terms used, including bias, unfairness and discrimination;
  - Define and share statistical definitions of bias, unfairness and discrimination evaluated by the auditing tool
- A tool should be developed to inform requirements for, and the choice between, statistical definitions for the context in which they are used

Invitation to collaborate

Based on the findings in this report, we invite leaders and champions of equality to co-develop and pilot an equality impact assessment with us.

See our first iteration of this equality impact assessment in Annex 1.

Regular auditing for equality, and taking steps to make appropriate adjustments where inequality is identified, are required to avoid breach of the Equality Act where AI systems are used to determine access to work.

Context

Widespread adoption of artificial intelligence (AI) is transforming work and lives across the UK.[1] 98% of Fortune 500 companies use AI or data driven systems at some stage of hiring[2]. The coronavirus crisis is accelerating the adoption of such systems to recruit, evaluate, track and onboard employees.[3] Employers are 'panic-buying' automated onboarding and monitoring systems. Amazon, for instance, recently used data-driven technology to on-board 1,700 staff in a day.[4] Serco has cut the time it takes to hire a worker from 4 weeks to 4 days.[5]

The Institute for the Future of Work's research suggests that AI tends to be used in three primary areas in the workplace: hiring, management and performance review.[6] As our economy adjusts to the shock of the crisis, and several cycles of social distancing and self-isolation over the coming months or years,[7] coronavirus will accelerate several prominent future of work trends. This includes the increasingly pervasive use of automated hiring systems. The Institute for the Future of Work anticipates more frequent transitions of workers between firms and sectors, as some stall, such as hospitality and traditional transport, whilst others grow, for example delivery services, on-line retail and technology.[8]

At a time of such profound change, equality matters more than ever. Human decisions about how AI systems are designed and deployed will shape access to and experience of work for generations to come. There is strong evidence that promoting equality is economically beneficial, as well as socially just.

At a time of such profound change, equality matters more than ever. Human decisions about how AI systems are designed and deployed will shape access to and experience of work for generations to come. There is strong evidence that promoting equality is economically beneficial, as well as socially just. Research has found that as much as two fifths of productivity growth since 1960 is the result of reducing barriers to women and BAME men;[9] and that promoting equality in the workplace would expose children in more families to factors that drive innovation, quadrupling the total number of inventors.[10] The equality agenda is becoming 'mainstream' within the business community.[11]

Statistical decision-making systems, using assumptions based on big data to predict future behaviour of individuals, are becoming ubiquitous in our economy, society, and government. So decision-makers must embed thinking about the implications for equality when designing and deploying these systems, in addition to considering fairness, accountability, sustainability and transparency (the 'FAST' principles) and data protection principles.[12] People are also more likely to accept the widespread use of AI if it adheres to the rule of law, and reflects common values of justice and equity.[13]

Comprehensive evaluation of AI systems is a process that includes auditing the consequences of correlations being drawn by AI systems, documenting these, and evaluating the differential impact of AI systems upon people with different characteristics across time. This process must ensure proper regard is given to *understanding* different types of impact on equality, offer frameworks for reasoning about trade-offs between those different types of impact, and should point to steps that can be taken to *mitigate* inequalities identified. Moreover, decision-makers should be encouraged to *actively promote* equality, even where doing so exceeds the strict requirements of equality law.

Our report proceeds as follows.

We demonstrate the case for evaluating the impact of AI on equality.

We review the different technical, statistical and necessarily narrow approaches to defining bias and fairness in AI which are often used in auditing tools.

We outline the inadequacies of these definitions in addressing the risks to equality of machine learning hiring systems, and present the expectations of UK equality law.

We review of existing tools for auditing AI systems in hiring, evaluating the strengths and limitations of each.

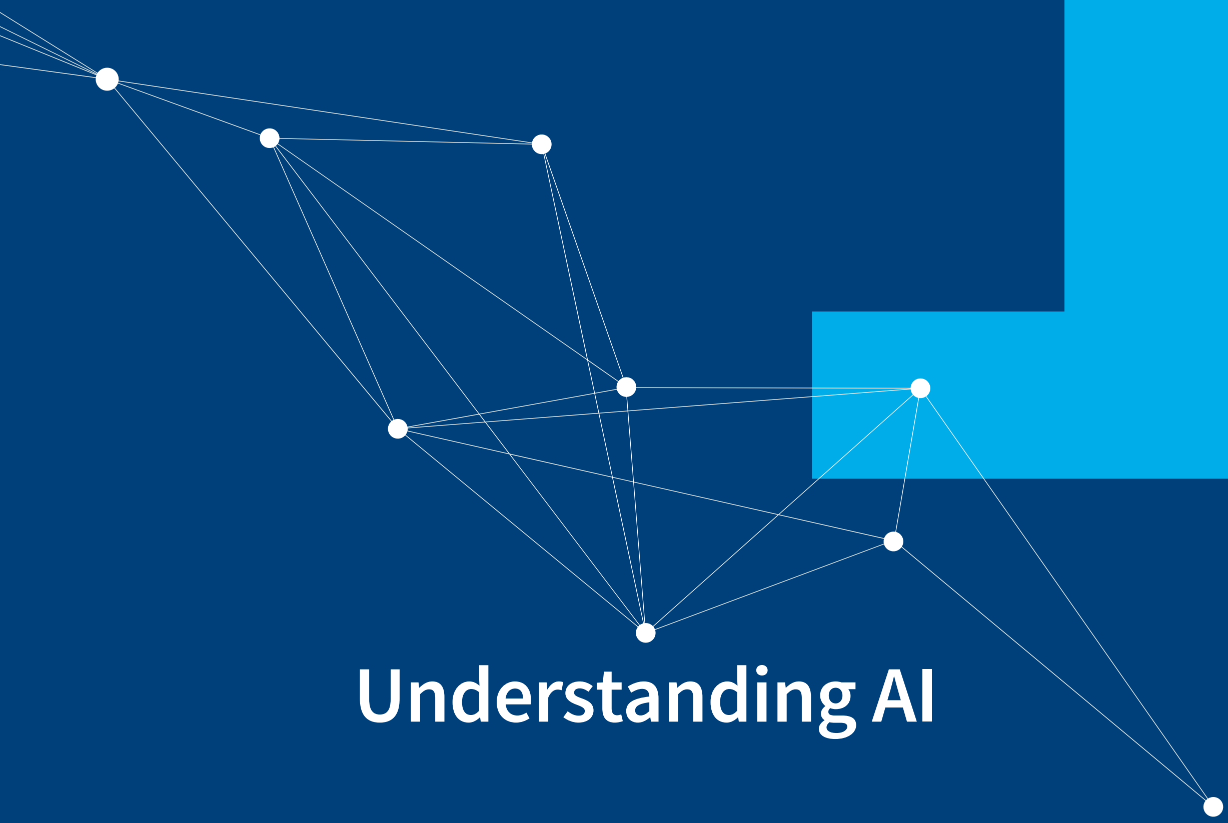
We outline how technical auditing should fit within a broader process of equality impact assessment (EIA).

Finally, we identify directions for future research, policy and legal development, focussing on where we think the stakes are highest.

Throughout this report, we draw on a strong body of work in the UK on AI governance and algorithmic decision-making. This includes draft guidance on an AI auditing framework published by the Information Commissioner's Office (ICO), several robust committee reports from the Houses of Commons and Lords, and reports by the Centre for Data Ethics and Innovation (CDEI), ADA and the Turing Institute.[14] This strong and growing body of work covers a range of approaches to making algorithmic systems compliant with the FAST and data protection principles. However, this work has not on the whole explored the impacts and implications of use of AI on *equality*. Reference is sometimes made to the prevention of discrimination, but what this means and how it can be done is not stated.[15] Further, discussion of legal requirements tends to focus on data protection law, rather than equality law. This means the growing challenges AI poses for structural inequality can be obscured, and the role of human decision makers minimised.[16] Our report begins to fill these gaps.

Discussion of legal requirements tends to focus on data protection law, rather than equality law. This means the growing challenges AI poses for structural inequality can be obscured, and the role of human decision makers minimised.

1



Understanding AI

Section 1

Understanding AI

> ‘ML is often presented in a way that suggests neutrality: an algorithm of independent capability with more processing power and less emotion than human actors. The reality is far messier. ML is a set of techniques designed by a human which addresses a problem defined by a human, trained on data-sets which usually encode the structures, opportunities and disadvantages of a very human landscape’.[17]

Artificial intelligence (AI) is a scientific field rather than a particular technology, which holds within it contested perspectives as to how it should function. Machine learning is a category of AI in which computers ‘learn’ from data how to accurately perform specific tasks – such as targeting potential applicants for a job. Instead of being explicitly programmed to follow a set of rules, machine learning (ML) systems learn how to accurately estimate an outcome from quantitative data sets, which are used to train and evaluate performance of a model over time.

The ML system will identify patterns from correlations between characteristics in data which match what has been defined as a ‘successful’ outcome in the past to predict for ‘success’ in future. Because ML systems use data about the past to inform decisions that shape the future, they can reproduce existing patterns encoded in training data, including historic inequalities among social groups: races, genders, classes, and geographic regions. In this sense, their interpretation as offering neutral and absolute insight is often misplaced[18].

ML systems may identify statistical correlations from a range of data points which no human mind has consciously identified as ‘relevant,’ continually absorbing new information and seeking new correlations as they learn. This renders any snapshot in time inadequate in understanding the risks posed to equality. In a broad range of contexts, ML can replicate, and potentially exacerbate, past patterns of bias, discrimination and inequality, reinforcing established patterns and projecting them into the future on an unprecedented scale.[19]

Section 1
Understanding AI

Machine learning systems are evaluated by how satisfactorily they appear to predict future 'success'. But because they are not human, they cannot critically evaluate past definitions of 'success' against which this future performance becomes a benchmark. This sits outside of their task and remit, and instead requires human judgement.

Consider a simple example: a machine learning model that estimates the probability that someone will click on a job advert online. The model predicts the probability that someone will click on a particular job advert. If the click behaviour of people online is stereotyped in gendered ways, for instance, men tending to click on adverts for 'NHS hospital manager' and women for 'social care worker,' then a machine learning model trained to estimate click probability will show men more adverts for managerial jobs and women adverts for caring jobs. If there are persistent inequalities in the incomes attached to these jobs, then this machine learning model will entrench patterns of gender-based inequality – potentially on a significant scale.[20] This is because ML models learn from past patterns of human behaviour, reflecting these patterns back to us. When an ML model's predictions are used in decision-making, this can entrench those patterns, creating a powerful feedback loop.

> Machine learning predictions are not predictions about a particular individual, but about groups who share certain features or patterns of behaviour. They leverage group patterns that most accurately predict relevant outcomes. A machine learning model learns that past patterns of behaviour are correlated among groups of similar gender, race, ethnicity, age, and so on.

> There is nothing inevitable about how AI shapes the future. How AI changes work, for instance, will depend on how governments and businesses design and deploy their AI systems. As AI is increasingly used to source, screen, select, and manage employees, policy makers and technical experts must develop systematic frameworks to evaluate a range of impacts on individuals, groups and society.

ML predictions are not predictions about a particular individual, but about groups who share certain features or patterns of behaviour. They leverage group patterns that most accurately predict relevant outcomes. A ML model learns that past patterns of behaviour are correlated among groups of similar gender, ethnicity, age, and so on. When the model replicates those patterns, that is not because it is *using* gender in its predictions but because it has 'learned' that gender, age or ethnicity shape click patterns. Any particular characteristic could be excluded as a variable from the model's training data and an accurate model would still learn these correlations.[21]

The "Equality Through Transition" paper, published by the Institute for the Future of Work, discusses how technical tools can replicate and reinforce social inequalities in greater detail. ML systems are designed by humans and reflect human choices. Humans must be accountable for how they design and deploy ML systems and how they evaluate the impact of those systems – particularly on equality.

This report uses the term AI for simplicity's sake, but almost all the tools we evaluate and examine primarily use machine learning.

Why evaluate the impact of AI on equality?

There is nothing inevitable about how AI shapes the future. How AI changes work, for instance, will depend on how governments and businesses design and deploy their AI systems. As AI is increasingly used to source, screen, select, and manage employees, policy makers and technical experts must develop systematic frameworks to evaluate a range of impacts of AI systems on individuals, groups and society. These frameworks will shape whether AI undermines, or furthers, the pursuit of equality at work. AI offers enormous opportunities to boost efficiency and productivity, but it also poses considerable risks that employers will unwittingly propagate patterns of bias, discrimination and inequality.

Existing approaches

Existing approaches address different aspects of evaluating the impact of AI on one type of equality. These are all necessary components of a holistic approach to AI equality impact assessments, but they are not sufficient. The narrowest and most focused component is to execute a technical and statistical audit of bias and fairness within an AI system, such as those we analyse in this report. A broader approach should also ensure compliance with legal requirements, such as the ICO's guidance on auditing which includes advice on when and how businesses should undertake a Data Protection Impact Assessment ('DPIA'). Beyond technical and legal components, impact assessments may also evaluate how AI systems impact the FAST principles, and other relevant principles and guidance.[22] Existing frameworks that focus specifically on risk assessment may also overlap with evaluating impacts on equality, including human rights, environmental and privacy assessments.[23]

Why go further?

The Institute for the Future of Work proposes an additional principle to guide evaluations of the impact of AI systems on individuals, groups and society: 'equality', as we define below. We think that this additional principle would enrich existing approaches is essential to guide both technical auditing tools and to wider impact assessments of AI systems.

We propose the integration of an improved technical auditing tool, which pays closer attention to equality, into a broader framework for evaluating a range of impacts on equality within equality impact assessments (EIAs). EIAs should be a critical part of effective regimes for governance, best practice and oversight as the use of AI becomes ever more widespread. This will build public and workforce trust in AI and contribute to building a future of better work and, in turn, a fairer society.[24]

Equality should be recognised as a guiding principle in the design and deployment of AI and auditing systems, alongside Fairness, Accountability, Sustainability and Transparency.

2

Auditing: Statistical definitions of bias and fairness

Section 2

Auditing: Statistical definitions of bias and fairness

Auditing something effectively – whether accounts, a policy, or an AI system – requires clarity about what you are auditing *for*. Clarity about definitions is critical to both designing AI systems and to evaluating those systems using auditing tools.

Computer scientists can only work effectively with corporate managers and policy makers if there is shared clarity about which definitions are being used at each stage.[25]

AI auditing tools apply particular definitions of bias and fairness, derived from and articulated by computer science research, to evaluate AI systems. There is often no 'correct' statistical definition of what constitutes an unbiased or fair AI system. This means auditing tools must be explicit and clear about which definitions they evaluate, what those definitions mean, and in what ways they are limited.

Consider the machine learning model introduced earlier, that estimates the probability that someone will click on a job advert online. This model predicts the probability that someone will click on a particular job advert. The model is trained on data about who has clicked on which job adverts in the past. Adverts the model predicts a particular user is most likely to click is shown to that user, whilst adverts the model predicts a particular user is not likely to click is shown to other users who are more likely to click on it. Let's call this model p(click).

The estimates of click probability this model produces make a significant difference to which users see which job adverts. This is exactly the kind of model used by companies like Google, Facebook, Amazon and others to help determine whether to show particular adverts to individual users. In this section we will use this example to illustrate how five popular definitions used in auditing are limited in their ability to respond to the various challenges set out by the Equality Act. While these definitions are commonly used it should be noted they are not comprehensive.[26]

1

Anti-classification

Anti-classification is perhaps the simplest definition of fairness – for that reason, it can also be misleading. Anti-classification holds that protected attributes – such as race or gender – and their close proxies should not be explicitly used to make decisions. In practice this is interpreted to mean protected variables are removed from training data sets and ML models.

Anti-classification stems from the idea that decisions should be "colour-blind" or "gender-blind". Its widespread grip on current approaches in computer science is driven by narrow interpretations of discrimination law, particularly in the US.[27] However, anti-classification will fail to secure fairness and may often undermine it.

First, it is often unclear which variables should be removed under anti-classification, since "close proxies" is a vague concept. The US Department of Housing and Urban Development (HUD) recently ran into this problem when defining discrimination in AI.[28] As outlined above past click behaviour for job adverts could generate problematic if predictable, gender bias. However, as ML systems can take in a wide range of datapoints and relationships, less obvious considerations can be taken into account. For instance, if past 'successful' staff at a company were all found to like skiing related Facebook pages, or shop at Ocado, or live in more affluent postcodes, those with similar (affluent) characteristics would become preferred candidates.

Second, removing protected features and close proxies in the learning process can make models both less accurate *and* less fair. Anti-classification blinds ML models rather than ensuring they do not replicate patterns of inequality. For instance, removing gender from the training dataset of the p(click) model will not prevent the model from replicating the stereotyped click behaviour of users. In fact, removing gender may reduce the performance of the model overall, exacerbating disparities in the job adverts shown to men and women. It may be better to ensure ML models make accurate predictions, rather than blinding them, as part of an overall decision-making system that respects the requirements of discrimination law and the pursuit of equality. The machine learning community has begun to coalesce around the view that unless legally required, models should not be arbitrarily blinded to protected traits.[29]

2

Calibration

The intuitive idea behind calibration is simple. Calibration means that people who have a particular probability or risk score can be understood to have that probability or risk score regardless of their gender, ethnic group, or other demographic attributes.

If a model identifies a set of people as having probability – '*p*' – of clicking on a job advert, the model is well-calibrated if approximately *p* fraction of people do in fact click on the advert. Models should be well-calibrated across different protected groups. This is equivalent to requiring that for a set of people with probability *p* of clicking on an advert, outcomes should be independent of protected attributes.

So to continue our earlier example, a model would be well calibrated if it estimated correctly – as confirmed by the training dataset – that women had a higher probability than men to click on the 'social care worker' advert than 'hospital manager' advert.

Calibration is desirable in almost all circumstances. However, calibration should be thought of more as a measure of good practice in machine learning than as a guarantor of fairness.[30]

3

False positive and false negative rates

In statistics, when performing multiple comparisons, a false positive ratio is the probability of falsely rejecting the null hypothesis for a particular test. Where a system assigns positive or negative scores to people, false positive (FPR) and false negative rates (FNR) can be used to measure how often the model incorrectly assigns a positive or negative score across protected groups. One common definition of bias and unfairness holds that ML models should not systematically make mistakes, in either the positive or the negative direction, across protected groups.

To explore this with our example, a training data set which presented evidence that 20% of female viewers of the care job advert clicked on it, whereas 3% of male viewers did, may generate the null hypothesis that women were more likely to respond to the advert. Women, or a proxy representing them as a classified group, would then be assigned a positive score. If this was found to be 'correct' in implementation, with more women clicking on the advert than men, the nul hypothesis would be accepted, leading to a learned rule. In turn, the rate would not be deemed 'false' by an auditing tool using this definition of fairness.

Whilst FPR and FNR are intuitive definitions of fairness, their relevance depends heavily on context.[31] In particular, the three definitions of bias and fairness cannot all be achieved simultaneously. If an outcome is distributed unevenly across two social groups, a model which predicts that outcome cannot be well-calibrated *and* have equal false positive rates *and* have equal false positive rates across a protected group. In our example, if the probability a user clicks on a particular advert differs across men and women, a model that predicts click probability cannot both be well calibrated and have equal FPR and FNR. When data encodes patterns of inequality – as it usually does in our own unjust world – there are conflicts between different statistical definitions of fairness.[32]

4

Demographic parity (The 4/5ths rule)

Demographic parity holds that a model must assign an average probability that is equal (or in some fixed proportion) across two groups. This means that p(click) should produce estimates of click probability that are equal (or in some fixed proportion) across different genders or racial groups.

In the US, a form of demographic parity is required by the Equal Employment Opportunity Commission (EEOC) according to the 4/5ths (or P%) rule. The rule states that the ratio of probability of selection of the lowest probability group to the highest probability group should not be less than 80%. For instance, if 60% of male applicants are invited for interview, no less than 48% of women should be invited for interview.[33]

Demographic parity is a crude measure of fairness. Often data reflects real differences between groups of people, differences that are not natural or inevitable, but which are produced by historic patterns of inequality and injustice. This might include the stereotyped click behaviour of users encoded within p(click)'s training data. Artificially imposing a consistent statistical pattern on these differences can obscure policy questions about how to make decisions in the context of social and economic inequalities.[34]

Section 2
Auditing: Statistical definitions of bias and fairness

5

Counterfactual fairness

Counterfactual fairness also corresponds closely to the sort of analysis which is used in equality law analysis, i.e. identifying if a particular protected characteristic played an unjustified causal role in decision making about an individual by asking what would have happened 'but for' the person having that characteristic.[35] [It holds that given a classification for a particular person, had that person's sensitive feature(s) been different, then all else equal, the classification would have been the same. If the model estimates the probability *p* that a woman clicks on an advert, if her gender is flipped, the model should produce the same probability *p*.]

Again, whilst intuitive on the face of it, the benefit and consequences of applying counterfactual fairness depends heavily on context.[36] For instance, requiring that p(click) produces the same click probability to a user if their gender were flipped means artificially ignoring genuine differences in the correlation of click behaviour with gender. It might, for instance, result in Google, Facebook or Amazon showing job adverts to people they have no interest in applying for.

There is often no 'correct' statistical definition of what constitutes an unbiased or fair AI system. This means auditing tools must be explicit and clear about which definitions they evaluate, what those definitions mean, and in what ways they are limited.

Technical definitions in review

These technical definitions are useful indicators for evaluating the impact of an AI system on equality, but they are not sufficient. They measure whether an AI system [respects or violates particular statistical patterns, but they have no bearing on whether those patterns are appropriate or justified, and if they are not, why they not]. This means they can measure whether a ML system creates statistically different outcomes for members of particular protected groups, but cannot show whether this is 'fair' or 'discriminatory' without identifying the causes of the statistical disparities, is beyond their design capability, thus instead requires human analysis.

Consider an example. P(click) might be well-calibrated, respect the anti-classification definition of fairness, and not violate the 4/5ths rule. And yet, it might still reinforce patterns of inequality on an enormous scale, consistently showing job adverts with lower average salaries to women than men. This is not because of unequal FPR or because the model violates counter-factual fairness, the other two definitions of bias and unfairness. It is because the model reflects back to us a persistent pattern of social inequality, produced by the long history of excluding women from the labour market.

Narrow definitions of bias and fairness can help decision-makers understand what systems are doing, but they offer no guidance about how those findings should be interpreted given that data encodes historic patterns of injustice.

We argue this is what an audit tool serving the purpose of equality should be doing: identifying different patterns of inequality and exposing the correlations which are being drawn, so that a decision-maker can assess how these patterns arise and what can be done to mitigate them. This might, for example, involve adjusting the training data fed into the ML model or the correlations it is and is not programmed to identify.

3

Auditing equality: UK legal requirements

Section 3

Auditing equality: UK legal requirements

There are long-standing disagreements about what is meant by 'equality' and whether what should be promoted is a 'fair' (or equal) outcome, or an equal opportunity, and whether 'equality' is measured on an individual or group basis.

The UK legal framework, set out in the Equality Act 2010 ('EA'), uses several definitions of 'equality' and how to promote them. It does not permit individuals to be treated differently because of a particular 'protected characteristic', which includes age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, sexual orientation without express statutory authority. This is the prohibition of direct discrimination. It does not permit the use of provisions, criteria or practises (PCPs) which have different adverse effects on members of groups with particular protected characteristics unless the use of such PCPs can be shown to be a proportionate means of achieving a legitimate aim. This is the prohibition of indirect discrimination. It requires certain 'reasonable adjustments' to be made for disability. It also requires public authorities to give 'due regard' to the needs to eliminate unlawful discrimination, advance equality of opportunity and foster good relations between members of different groups.

This spectrum of requirements under the EA demonstrate that the purpose of the EA is to produce *'better outcomes for those who experience disadvantage,'* as the Explanatory Notes reiterate.[37] The Equality and Human Rights Commission explains that the focus of the EA is to *'protect the rights of individuals and advance equality of opportunity for all.'* The Commission goes on to explain that the EA includes wider duties and mechanisms which aim to *positively advance* equality, including reducing certain types of equality of *outcome*. Taken together, these duties suggest that the EA is designed to tackle systemic or 'structural' inequalities, as much as protect individual rights. This is an important component of the UK approach in an age of AI since, as we have seen, ML Systems can replicate inequalities that do not conform to narrow definitions of discrimination, and can identify groups with common traits not classified as protected characteristics, making decisions over time which could disadvantage and or exclude them from the labour market at scale.

In addition, as the Institute for the Future of Work describes in Equality Through Transition, the impacts of automation and AI affect equality at a systems and firm, as well as individual, level.

Table 1: Auditing and evaluation of AI systems for equality

Definition	Remit	Audience	Is it in force?
Direct discrimination (s13 EA)	Individual rights to challenge less favourable treatment which occurs because of protected characteristic	Public and private	Yes
Indirect discrimination (s19 EA)	Individual and group rights to challenge a practice which puts a group of people who share a protected characteristic at a disadvantage	Public and private	Yes
Making adjustments (s20 EA)	Positive duties to make reasonable adjustments and offer extra support to avoid substantial disadvantage for disabled people	Public and private	Yes
Equality duty (Part 11, especially s 149)	Positive duties to advance equality of opportunity and foster good relations between individuals and groups when making decisions	Public	Yes
Reduction of equalities of outcome (s 1 EA)	Duty to consider means to reduce known inequalities of outcome resulting from socio-economic disadvantage when making decisions	Public	Not in England and Wales; only in Scotland

Five of the most relevant definitions and mechanisms to steer auditing and evaluation of AI systems for equality are shown in Table 1.

The ways in which the Equality Act may apply to the use of a ML system, including in hiring, is mostly untested. This means the scope and content of requirements on employers and others have yet to be fleshed out.

It may be difficult for a person to whom an AI system has been applied to know, or show, what the adverse effect of this has been on them, and what the link is with a particular protected characteristic. Our research has found that ML hiring systems can take into account factors such as a candidates' post code, educational background or even voice which may be proxies for socio-economic background as much as proxies for a protected group.

If the disparity of effect is established using statistics, an employer may struggle to justify this if it cannot explain correlations drawn by its AI system. Employers are required to show a 'real business need' and that use of the AI system is 'appropriate' to achieve a 'legitimate' aim in order to avoid a court drawing an 'adverse inference' of indirect discrimination. But, it is currently unknown how case law on causal reasoning will apply to the 'black box' of an algorithm when its designer may not understand the learned basis by which a statistical pattern has been discerned.[38]

Some commentators suggest that this may well mean that most uses of AI would constitute unlawful indirect discrimination, which is legally problematic for employers. An employer who has not predicted and sought to mitigate the impact of use of AI on disabled applicants or employees is likely to be in breach of the duty to make reasonable adjustments to its practices to avoid unlawful disability discrimination. And a public body which has not considered these issues or considered how it could advance equality may also be in breach of the public sector equality duty.

In these circumstances, there are sound business, technical, legal and policy reasons why employers should aim to *exceed* the strict requirements of the law. Exercising caution in this way, and championing good conduct, will reduce the risks of a claim or finding under the Equality Act and support the development of AI systems that promote equality. It will also help establish new norms in best practice and build trust in AI, at a critical time in its development and use. Our approach, which suggests use of an equality impact assessment (EIA) to assess the effect of use of AI, aims to help employers achieve those goals.

In turn, ‘equality’ should be defined as more than the absence of direct or indirect discrimination. In line with the fundamental purpose of the EA, the definition of ‘equality’ should capture each of the five models identified in this section. Specifically, evaluation of impacts on equality should extend to assessment of the risks and compliance with the five definitions above.
AI systems should conform to the requirements of law in the jurisdiction in which they operate. However, as we have seen many technical definitions used in auditing are not equipped to achieve this. In turn, it is risky and problematic for regulation to be adapted to become translatable to enforcement through computer systems. To move forward, it is necessary to have a rigorous debate about responsibilities, policy and interpretations of the law.

There are sound business, technical, legal and policy reasons why employers should aim to *exceed* the strict requirements of the law. Exercising caution in this way, and championing good conduct, will reduce the risks of a claim or finding under the Equality Act and support the development of AI systems that promote equality.

Recommendation

We recommend that large employers, i.e. those with more than 250 employees, which use AI tools in their hiring practices, should demonstrate leadership by conducting equality impact assessments as if sections 1 and Part 11 EA applied to them.

4

Existing auditing tools

Section 4

Existing auditing tools

This report reviews existing auditing tools across a range of AI applications in the recruitment process. Our review is consistent with literature reviews which identify 10 underlying tasks that can be audited in relation to recruitment.[39]

The 10 underlying tasks
Vacancy prediction software
Job description optimisation software
Targeted job advertising optimisation
Multi-database candidate sourcing
CV screening software
AI-powered psychometric testing
Video screening software
AI-powered background checking
Employer branding monitoring
Candidate engagement chatbot/customer relationship management

These ten tasks are divided across four key stages of hiring: sourcing, screening, interviewing, and selection. Automated Hiring Systems (AHSs) are software-based tools that support the hiring process. Some AHSs support all four stages of the hiring process whilst others support only a subset of stages. Each stage involves different decisions and data, and requires different approaches to identifying and mitigating bias, discrimination, and the reproduction of inequalities (Table 2).

Sourcing, the first stage, involves soliciting candidates, inviting applications, and advertising for the position. Tools like Textio evaluate whether the language employers use to advertise positions and solicit applications unwittingly favours particular genders or ethnic minorities. Screening, the second stage, involves evaluating whether candidates should be invited to interview, using information about their CV, and their job application. Auditing tools like Ideal aim to ensure that systems used to invite candidates for interview do not use sensitive data and respect legal definitions of statistical unfairness, such as the 4/5ths rule in the US.

Interviewing and selection, the third and fourth stages, involve using all available material to evaluate the likely performance of candidates. A variety of tools exist to evaluate the use of AI systems in these stages of the hiring process, including tools such as What-If, Aequitas, or AI Fairness 360.

Table 2: The four stages of the hiring process

Stage	Decisions	Data	Example of bias-reducing approach
1. Sourcing	Confirm opening Description and criteria Screen Reach out	Job advertisement and selection criteria Social media (e.g. LinkedIn), proprietary profiles (i.e. candidate platform), listserv, referral database	Opening, selection criteria, and advertisement are derived and described in neutral way Propensity of outreach is independent of sensitive attribute Presence is equal across attributes
2. Screening	Pass for interview Match to job	All data plus CV/resumé, candidate response, candidate performance on test, internal performance data	Pass (outcome) is independent of sensitive attribute Model has no knowledge of sensitive attribute Match is independent of sensitive attribute
3. Interviewing	Pass/Fail Classify performance	All data above plus question responses, visual/auditory data	Model representation is independent of sensitive attributep
4. Selection	Hire Compensate Train Match to job	All above plus interviewer evaluations	Outcomes, in addition to current workforce, meets a predefined diversity standard

Some Automated Hiring Systems support all four stages of the hiring process whilst others support only a subset of stages. Each stage involves different decisions and data, and requires different approaches to identifying and mitigating bias, discrimination, and the reproduction of inequalities.

AI can be used in most of these four stages: to source candidates, for instance, by using AI in job advertisement; in screening candidates, for instance, in determining which candidates to invite for interview; and to inform selections, for instance to predict future job performance on the basis of sales, personality traits, job tenure, and other metrics. However, few employers have automated the entire hiring process, particularly the interviewing and selection stages. AI systems used in any one of these four stages can introduce bias, unfairness, and discrimination. For this reason many auditing tools are designed to present findings which are easy for a human to interpret. In the context of ML systems, this is called ‘interpretability’.

Table 3: The four stages of building an AI system

Stage	Challenges	Term for mitigation	Address at this stage if...
1. Data	The data entering the model reflects biases you don't want the model to replicate (i.e. patterns of inequality across gender, race, performance). The training data is not representative of the broader population.	*Pre-processing*	You can modify the data; it is easy to modify the data; the data has a clear representation of sensitive attributes (sometimes true, sometimes not); data is not streaming; your pre-processing procedure removes bias.
2. Model	The model itself has been written in such a way to create bias.	*In-processing*	You can find bias in your model (usually hard); you can extend your model to track and output fairness measures.
3. Optimisation criteria	The function a model learns to optimize reflects a bias, by replicating unfair or discriminatory patterns. The outcome a model is trained to predict has different meanings across different social groups, or is distributed unevenly.	*In-processing*	You can write a bias-adjusting optimisation criterion (sometimes easy, sometimes hard).
4. Predictions	The predictions have not been adjusted after-the-fact to remove bias.	*Post-processing*	You can intervene on predictions before they are used; like in *pre-processing*, it is clear how to intervene to ensure fairness.

Building an AI system – in most cases using machine learning – is a process. Designing and deploying AI systems involves a series of choices made by engineers or data scientists, embedded within particular organizations, with different incentives, policies, and laws.

These choices are made at four particular stages of building an AI system: assembling the data on which the system is trained, selecting the model, choosing the objective the system will optimize, and evaluating the predictions it produces. Bias, unfairness and discrimination can be introduced – or more often replicated, if it exists in the data – in each of these four stages (Table 3).

Section 4
Existing auditing tools

Hiring: Overview of auditing tools

With this hiring process and the several possible sources of bias, unfairness and discrimination in mind, we analysed a range of existing auditing tools for AI in the workplace. We followed a three stage search for tools.

First, we considered AHS full solutions and tools that had been reviewed in literature or had self-published research.

Second, we expanded to include commercial and open-source tools for interpretability tools focused on fairness and bias. We sourced commercial tools via their websites and open-source tools largely by searching for active and well-followed repositories on fairness on GitHub.

Last, we incorporated a select few that we noticed are increasingly used in applied research around interpretability and fairness. While we believe this list covers the major tools, it is not comprehensive, nor is it long; the field is just emerging, and so are the tools. Here, we mostly focus on detection[40].

We explored each existing tool, analysing its core features and applications. We focused mostly on publicly available information, whilst obtaining some information via private channels where necessary. We evaluated what definition(s) of bias or unfairness the auditing system applies, where that information is clearly noted and publicly available (Table 4).

In the context of machine learning systems, interpretability is the ability to explain or to present in terms which are understandable to a human.

Table 4: Overview of auditing tools

Tool	Type	Description	Unique value	Definition(s) used
audit-AI	Commercial AHS, Open-source	Developed by pymetrics internally, and then open sourced, audit-AI is a tool for *detecting* bias in a machine learning algorithm.	Integrated into a commercial end-to-end AHS platform.	Demographic parity (4/5ths rule); anti-classification; counter-factual fairness.
Ideal	Commercial AHS	An intelligent AHS that can test for the 4/5ths rule.	Specific application of US-relevant fairness criterion.	Demographic parity (4/5ths rule); anti-classification.
Textio	Commercial AHS	An intelligent writing assistant that can detect bias in language (e.g. gendered language).	Operates at the very beginning of the pipeline (sourcing).	*Method not explained*
MeVitae	Commercial AHS	An intelligent CV pre-processing tool that removes potentially bias-inducing information from the CV.	Operates at the very beginning of the pipeline (sourcing).	Demographic parity (4/5ths rule); anti-classification.
Fairness Flow	Commercial, Internal	Facebook's internal, and still secret, bias detection and mitigation tool.	Opportunity to be deployed at scale in one of the largest consumer-facing products.	*Method not explained*
Aequitas	Open-source	Developed at the University of Chicago, Aequitas is a detection suite (like Fairness Comparison, FairTest, FairLearn, and FairML). A unique feature is a "Fairness tree" that allows researchers to find the correct fairness metric for their task by answering simple questions.	A "Fairness tree" that helps to choose the right definition.	Calibration; demographic parity; anti-classification; FPR and FNR; equal opportunity.
AI Fairness 360	Open-source	A suite from IBM to detect and mitigate bias in the pre- and post-processing stages. Major advantage is the number of implemented metrics and processing methods. Compared to other bias detection suites, AIF360 also enables mitigation.	A robust library of many different definitions of fairness.	Demographic parity; anti-classification; FPR and FNR.
InterpretML	Open-source	A Microsoft developed toolkit to interpret machine learning models, including some metrics for bias. Note that interpretability metrics may differ from bias metrics.	A robust data and model exploration tool for data scientists.	Calibration; demographic parity; anti-classification; equal opportunity.
What-If	Open-source	A Google developed toolkit for observation-level counterfactual fairness evaluation (among other non-bias related objectives).	A "Fairness tree" that helps to choose the right definition. A clear and simple user interface designed at policy as well as technical audience.	Calibration; demographic parity; anti-classification; FPR and FNR; equal opportunity.

continued

Table 4: Overview of auditing tools *continued*

Tool	Type	Description	Unique value	Definition(s) used
FairML	Open-source	A popular fairness evaluation software library developed by a relatively well-known fairness in machine learning researcher. Detects variable contribution to model decisions.	*Not discerned*	Evaluates statistical importance of model's inputs.
FairTest	Open-source	Developed at Columbia university, FairTest is a tool for detecting subgroup fairness in an algorithm. Subgroup analysis finds bias on intersectional sensitive groupings.	*Not discerned*	Demographic parity; anti-classification; FPR and FNR.
SHAP	Open-source	A recently-developed, very popular tool for interpreting the effect of a particular variable on a decision outcome. Not a bias-detection tool but can be used as such by implementor. (Similarly, there are hundreds of papers on different methods for interpretability.)	Uses a recently-popular, well-understood interpretability criterion.	Focused on explainability measures including local linear explanation model (LIME) and Shapley Additive Explanations (SHAP).
Fairness Comparison	Open-source	Developed at Haverford College, Fairness Comparison is a toolkit for comparing interventional/counterfactual fairness of different models on benchmark datasets.	An easy benchmarking tool to evaluate models *before* deploying on real-life data.	Counter-factual fairness; demographic parity; equalized odds.
FairLearn	Open-source	Another popular fairness detection suite focused on *harm* detection.	A fairness suite that focuses on harm.	Calibration (quality of service harm); allocation harms (FPR and FNR).
TensorFlow Fairness Indicators	Open Source	Part of Google's popular TensorFlow library, TensorFlow Fairness Indicators is a tool for automating the detection and visualization of disparities in fairness-relevant metrics across sensitive attributes.	Easy integration in industry-leading machine learning framework.	Statistical discrepancies in user-defined metrics (e.g. FPR, FNR, Accuracy).
ML-fairness-gym	Open Source	A Google-developed framework for simulating systems of agent-algorithm interaction in order to evaluate long-term, emergent outcomes of algorithms, especially with respect to fairness.	Evaluate long-term impact on fairness.	User-defined
Fairness in Classification	Open Source	A single-purpose repository to implement fair logistic regression classifiers using a novel fairness definition.	Focused on mitigation for a single model.	Preferred Treatment/ Impact: Nash-equilibrium-like definition.

5

The limits of existing tools

Section 5

The limits of existing tools

Our analysis found there are a broad and growing range of auditing tools that can be used to detect different forms of bias and unfairness. Many of the commercial tools, which can be integrated into firms' existing HR systems, offer limited public information about how they define fairness.

The freely available public tools tend to use open-source software that anyone can run using Python, a general purpose high level programming language, many of which have been developed by companies like Google, Microsoft and IBM. These tools apply different statistical definitions of bias and fairness, most commonly those outlined in this report. Several offer clear explanations of what these definitions mean and how to apply them, such as Google's What-If.

However, there are several crucial limitations to existing tools. Some of these are about the definitions applied by the tools themselves, but most pertain to the limits of auditing without a broader framework for evaluating equality. Over the coming years, businesses, governments and regulators must work together to integrate frameworks, processes, policies, and auditing tools to evaluate the equality impacts of AI systems in holistic, systemic fashion. Auditing tools on their own are useful and often necessary but almost never sufficient to evaluate the implications of AI for equality.

1

Auditing tools are focused on US law and regulation

The first and most obvious limit is that because many AI auditing tools are produced in the US, they import US definitions of discrimination and use definitions of bias and fairness which most closely correspond to them. In particular, the 4/5ths rule required by the EEOC and the anti-classification definitions are present in almost all auditing tools.

The 4/5ths rule, which is a criterion of demographic parity, imposes an arbitrary proportional rule on social groups where there are often genuine differences between them in the underlying data. Far from serving as a useful way to detect discrimination, application of the 4/5ths rule in auditing tools may violate UK discrimination law, according to several UK legal experts.[41]

It is worth noting that divergences between discrimination law in the UK and the US are ultimately driven by different ways of defining what discrimination is, what kinds of discrimination are wrong and prohibited, and by divergent approaches to placing burdens on public and private sector organizations to mitigate it.[42] The UK and the US take different approaches to defining the obligations of private and public organizations to confront entrenched inequalities and to evaluate how their decision-making systems further or undermine that goal.[43]

Recommendation to developers

A tool should be developed to inform requirements for, and the choice between, statistical definitions for the context in which they are used

An improved technical auditing tool should be developed which pays close attention to the requirements and purpose of UK Equality Law

Recommendation to policy makers

CEOs and HR leaders need practical guidance on effective auditing and its wider role to promote legal compliance, good governance and best practice

Recommendation to business

To promote legal compliance, good governance and best practice, auditing must be aimed at exceeding the requirements of national equality legislation, as well as data protection and employment legislation

2

Auditing tools are often not explicit about how they define bias and fairness

There is no single, correct definition of bias and fairness. Computer scientists have reached broad agreement that imposing any single definition on AI systems used in different sectors and contexts is likely to be at best ineffective and more likely counter-productive. However, Corbett-Davies and Goel demonstrate that choosing one particular definition of bias and fairness can have unintentional consequences:

"... we show that all three of these fairness definitions suffer from significant statistical limitations. Requiring anti-classification or classification [demographic] *parity can, perversely, harm the very groups they were designed to protect; and calibration, though generally desirable, provides little guarantee that decisions are equitable."*[44]

This clarifies the challenge when auditing AI for equality. Auditing tools must be explicit and clear about how they define bias and fairness. They must identify the metrics and measurements they use. Where relevant, they must identify possible trade-offs among these particular definitions, and explain why those trade-offs are necessary.

Every auditing tool should state clearly, in plain prose and statistical terms, the different definitions of bias, fairness and equality used. They should also be clear about the sensitive attributes with respect to which they evaluate bias, fairness and equality.

Some auditing tools achieve this relatively successfully, such as Google's What-If and IBM's Fairness 360. Others do not. In particular, we found that many of the commercially available tools like Audit AI and Ideal offer limited public information about how they define sensitive attributes, bias or fairness. This limits the guarantees offered to job candidates or employees about the auditing of an employer's AI systems.

Over the coming months, the The Equality Task Force on Equality will publish further research and analysis which explores how organizations can evaluate and reason about the trade-offs between different forms of bias, fairness, and inequality in AI.

Recommendation to developers

Define the **sensitive attributes** with respect to which they evaluate fairness and bias;

Offer **definitions** of bias, unfairness or discrimination the auditing tool seeks to detect with respect to those sensitive attributes;

Offer **statistical definitions** of bias, unfairness or discrimination the auditing tool seeks to detect with respect to those sensitive attributes

3

Auditing tools are not widely adopted or available – and there's a capability gap

It is hard to evaluate which organizations are using AI auditing tools in hiring. However, it is likely that adoption is limited – in particular, limited to companies with the resources and technical capabilities to deploy these tools and to adjust their AI systems based on what they find.

This gap is likely to widen over time. In 2019, less than a third of CEOs who admit that they collect extensive data on their workforces personally feel that their companies use the data responsibly.[45] Without guidance and support, companies who fear what they might find after applying AI auditing tools may refrain from using those tools, further limiting technical capacities and increasing the capabilities gap. Proactive interventions and incentives will be required to broaden the adoption and use of AI auditing tools.

Recommendation to developers

Regular assessment, with appropriate adjustments, is required where AI systems are used to determine access to work

Before deploying AI systems to hire, companies should consult their workforce, or their affiliated union, to discuss equality

Recommendation to policy makers

Professional and industry standards for auditing tools, including auditing for equality, are urgently required to main high, consistent standards. We recommend that this initiative is led and coordinated by the CDEI

4

Auditing tools focus on detection but often overlook mitigation

Most of the auditing tools we analysed aim to detect bias or unfairness but do not seek to identify what steps an employer could use to identify *why* particular inequalities emerge and – to the extent that these are based on irrelevant characteristics or simply reproduce past patterns of disadvantage – to correct for, and mitigate them.

This is because technical auditing is itself focused on the detection of bias or unfairness, rather than deciding how best to address it or contextualise this information within a wider discussion of equality. However, those developing auditing tools should also develop methods and processes that will help human decision-makers reason about how to mitigate bias or unfairness detected in the auditing process. Tools could outline different courses of action, including possible trade-offs between them.

Auditing tools will never – and should not attempt to – automate the process of deciding how best to address bias, unfairness or persistent inequalities in outcomes produced by AI systems. Tools should, however, clarify the different possible alternatives and trade-offs.

Our snapshot analysis suggests that Google and IBM's tools may come closest to achieving this. In addition to offering explanations of different definitions of bias and unfairness, these tools allow AI auditors and designers to explore the trade-offs between those different definitions and the implications of adjusting an AI system to respect them.

Recommendation to developers

Companies should integrate technical auditing into a wider equality impact assessment

Recommendation to policy makers

Computer scientists and policy makers should work together to develop (i) a greater understanding of the risks presented by AI to patterns of systemic inequality over time (ii) develop approaches to addressing bias, fairness and promoting equality

Auditing must sit within a wider approach to evaluating the impact of AI systems on equality and work. New legal questions and ethical responsibilities presented by impacts which play out at an individual, firm and systems level must be considered

Over the coming years, businesses, governments and regulators must work together to integrate frameworks, processes, policies, and auditing tools to evaluate the equality impacts of AI systems in holistic, systemic fashion.

6

Equality impact assessments

Section 6

Equality impact assessments

As we have seen, technical auditing is an important but not sufficient component of evaluating the impact of an AI system on equality.

Even with all of the statistical definitions of fairness and bias auditing the (p)click model described earlier, this ML System could still reproduce and entrench the inequalities written into historic datasets. It is for this reason we argue wider, non-computational assessments are needed.

We have seen that:

- **Technical, statistical definitions used in auditing tools are inadequate to expose non-compliance with current legal duties outlined in the Equality Act**
- **AI systems continually evolve and may find new proxies for both protected characteristics, and other axes of discrimination, making a narrow or snapshot view inadequate**
- **The technical capabilities and limitations of AI systems point to the need for a deeper assessment of impacts on equalities of outcome, as well as opportunity**
- **There is a compelling ethical case to address the impacts of AI Systems on access to work, given impacts on equality of outcome over time**
- **There is a stated commitment by parts of the corporate community to advance equality**
- **There is a strong business case for having a more diversified workforce**

We therefore suggest that a more comprehensive evaluation of the impacts of AI on equality is necessary. We propose a framework which draws on and builds from existing impact assessments, and complements and reinforces relevant others, in particular Algorithmic Impact Assessments and Data Protection Impact Assessments,[46] but which is geared specifically towards advancing equality.

Our proposed Equality Impact Assessment aims to support and guide human evaluation of AI systems. For that reason, our EIA focuses on key human decision-making points in the design and deployment of an AI system: selection of the AI system; selection of the training data sets; selection of the outcome; and selection of the variables. We invite employers to voluntarily undertake this process. We will launch our consultation on the draft proposal for an EIA at Annex 1 on 29 May, the 50th Year Anniversary of the Equal Pay Act.

As proposed in our EIA, EIA outcomes should be part of collective agreement processes where employers recognise a trade union and initiate discussions about the assessment.

Our intention is that guided by an EIA, ML hiring systems work as they should: to promote equality, rather than embed inequality.[47]

7 Future work

Section 7

Future work

This report has considered the role of auditing in evaluating the impact on equality of AI systems used to determine access to work.

We have identified four particularly pressing areas for future work. Each broadens the focus from auditing to a more comprehensive framework for evaluating equality.

1

Legal codes and guidance

Legal Codes, from the EHRC and ICO in the UK, should provide detailed guidance on the application of the EA and GPDR, Data Protection Act. Guidance and statutory codes from our regulators have particular importance when clear interpretation and application is needed to inform design, as well as use.

Equality Impact Assessments should be commenced prior to the deployment of an AHS or AI system, enabling organizations to assess risks and evaluate potential impacts of their system, before it is deployed.

2

Industry and professional standards

Industry groups, sectoral regulators, workforce representatives and civil society should come together to determine best area-specific standards or practice in the light of legal codes and other guidance. These standards could extend to technical definitions aimed at both engineers tasked with designing systems and employers who must respect them.

We suggest the CDEI leads this initiative, supported by the Institute for the Future of Work and an advisory group.

3

Equality Impact Assessments

On top of industry and legal standards, Equality Impact Assessments (EIAs) should be developed across sectors. EIAs should be commenced prior to the deployment of an AHS system, enabling organizations to assess risks and evaluate potential impacts of their system, before it is deployed. Evaluation will then continue, extending to legal compliance and evaluation of actual impacts, and positive steps that can be taken at each key decision-making point.

4

Review of regulation

Our Equality Task Force will review the adequacy of regulation more generally in the light of new challenges brought by machine learning. Allen notes, “so far, no legislation has been passed that has been designed specifically to tackle discrimination in AI systems”.[48] A recent report from the Committee on Standards in Public Life similarly observed that “there is currently no bespoke regulatory guidance outlining what public bodes introducing AI systems need to do to comply with the Equality Act 2010.”[49] The report goes on to explicitly state that: “Government should remain open to the revision of anti-discrimination law if the current legal framework cannot answer these questions convincingly.”[50] Separately, the Council of Europe has recommended that nations review their legislative frameworks and policies as well as their own practices with respect to the procurement, design development and ongoing deployment of algorithmic systems.[51] We agree.

This review should consider the expectations of the Equality Act and challenges identified in this report. This may include the equality duty and provision to counter socio-economic disadvantage. As Dencick, Edwards, and Sanchez-Monedero write “asking providers of AHSs to attend to the dynamics of power in labour relations and society more broadly might seem unnecessarily burdensome, but by not recognising the broader functions of automation in shaping those dynamics... we run the risk of neutralising challenges in a way that actively facilitates discrimination and inequality under a banner of fairness.”[52] EIAs may inform review of legislation, and should not be seen as an alternative to it.[53]

> “Asking providers of Automated Hiring Systems to attend to the dynamics of power in labour relations and society more broadly might seem unnecessarily burdensome, but by not recognising the broader functions of automation in shaping those dynamics... we run the risk of neutralising challenges in a way that actively facilitates discrimination and inequality under a banner of fairness.”
>
> Javier Sanchez-Monedero, Lina Dencik, and Lilian Edward

> “Human decisions about how AI systems are designed and deployed will shape access to and experience of work for generations to come.

Institute for the Future of Work

Annex 1

Outline process for conducting an Equality Impact Assessment

1	**Preliminaries**	Should you undertake an Equality Impact Assessment?
2	**Purpose**	State the purpose of your AI System
3	**Assessing risk**	Conduct a risk assessment identifying potential risks to equality connected to use of your AI system
4	**Auditing**	Conduct a technical audit of your AI system
5	**Evaluation**	Evaluate the impacts on equality of your AI system and make necessary adjustments
6	**Explanation**	Explain and communicate the choices you have made

Annex 1

VERSION 1

Completing an Equality Impact Assessment

1 Preliminaries

Is your AI system used to determine access to work, or the fundamental terms and conditions of work?

IF YES

What are the key human decision-making stages in the deployment of your AI system?

Best practice
Draw a flow diagram of the business process in which the AI system is embedded, highlighting where the human decisions are made.

Best practice
Does your Equal Opportunities Policy cover the use and impacts of AI in your organisation?

2 Purpose

Can you describe your purpose in deploying this AI system?

Can you describe how you decided on the scope of the work to be undertaken by the AI system, the technical model you chose to undertake this work, and any alternatives you considered.

Have you consulted all key stakeholders, including workers and unions who may be affected by deployment of the AI system?

Best practice
Have you considered the advantages of deploying an AI system to actively promote equality?

Best practice
Are the decisions you have made at each key stage in the flow chart (page 42) aligned with your purpose?

3 Assessing risk

Have you assessed the risks of adverse impacts of equality of opportunity and outcome on individuals and groups with shared protected characteristics?

What common and statistical definitions have you chosen to undertake your advance risk assessment? Can you explain why?

Are you satisfied that your assessment will maximize legal compliance and good governance?

What adjustments could you make to address or mitigate the risk of adverse impacts on the individuals and groups you have identified?

Using objective criteria, have you considered which adjustments you can reasonably make? Have you in fact made those adjustments? If not, why not?

Best practice
Have you identified risks to individuals or groups that share a socio-economic status?

Best practice
Have you considered the likelihood and severity of the risks you have identified?

Best practice
Have you considered whether you can make an adjustment to the AI system, its use, or any other matter which could actively promote equality between the individuals and groups you have identified?

Best practice
Have you taken into account relevant industry standards or guides, as well as regulatory ones?

Annex 1

VERSION 1

Completing an Equality Impact Assessment *continued*

4 Auditing

Have you selected an auditing tool which can identify impacts on equality of opportunity and outcome between the individuals and groups you have identified?

Which tool/method/definitions have you chosen and why?

Can you articulate the strengths and limitations of the auditing tool you have deployed?

Using your tool, can you discern any actual adverse impacts on equality of opportunity and outcome for the individuals a for the individuals and groups you have identified?

Have you sought review of your EIA from an independent third party?

Can you demonstrate legal compliance?

Best practice
Does your technical audit point you to adjustments you could make at any key decision making stage in the flow chart (page 42)?

Best practice
Have you also considered how insights from your technical audit may enable you to contribute to advancing equality between the individuals and groups?

Best practice
Can you extend you audit to consider individuals and groups with an identified socio-economic disadvantage?

5 Evaluation

Have you integrated the results of a technical audit into a evaluation of the impacts of your AI system on equality of opportunity and outcome for relevant individuals and groups?

Using objective criteria, have you considered which adjustments you can reasonably make? Have you in fact made those adjustments? If not, why not?

Have you consulted all key stakeholders, including workers and unions who may be affected by deployment of the AI system?

Have you identified adjustments you could make to address or mitigate adverse impacts on equality of opportunity and outcome for relevant individuals and groups?

Have you sought review of your EIA from an independent third party?

Can you demonstrate legal compliance?

Best practice
Have you taken into account any relevant industry standards or guides, as well as regulatory ones?

Best practice
Have you considered how insights from your technical audit or any other part of this assessment may enable you to contribute to advancing equality between the individuals and groups?

Best practice
Can to extend your evaluation to individuals and groups with an identified socio-economic disadvantage?

Best practice
Have you considered adjustments to practice or policies outside use of your AI system?

Best practice
Have you considered adjustments you could make at each key decision-making stage?

Annex 1

VERSION 1

Completing an Equality Impact Assessment *continued*

6 Explanation

Can you explain the decisions you have made in the course of this assessment?

Have you publicly disclosed, or made publicly available, your statement of purpose?

Have you publicly disclosed, or made publicly available, relevant internal procedures and policies, including your equal opportunities policy and EIA plan?

Is there an accessible means and process for a person affected to seek and obtain an explanation for relevant decisions you have made?

Can you provide these explanations, including the adjustments you have made and an equality impact statement? Can you identify a person with overall responsibility for this EIA?

Have you consulted all key stakeholders, including workers and unions who may be affected by deployment of the AI system?

Have you explained how to apply for a human review of the decision, or part of it?

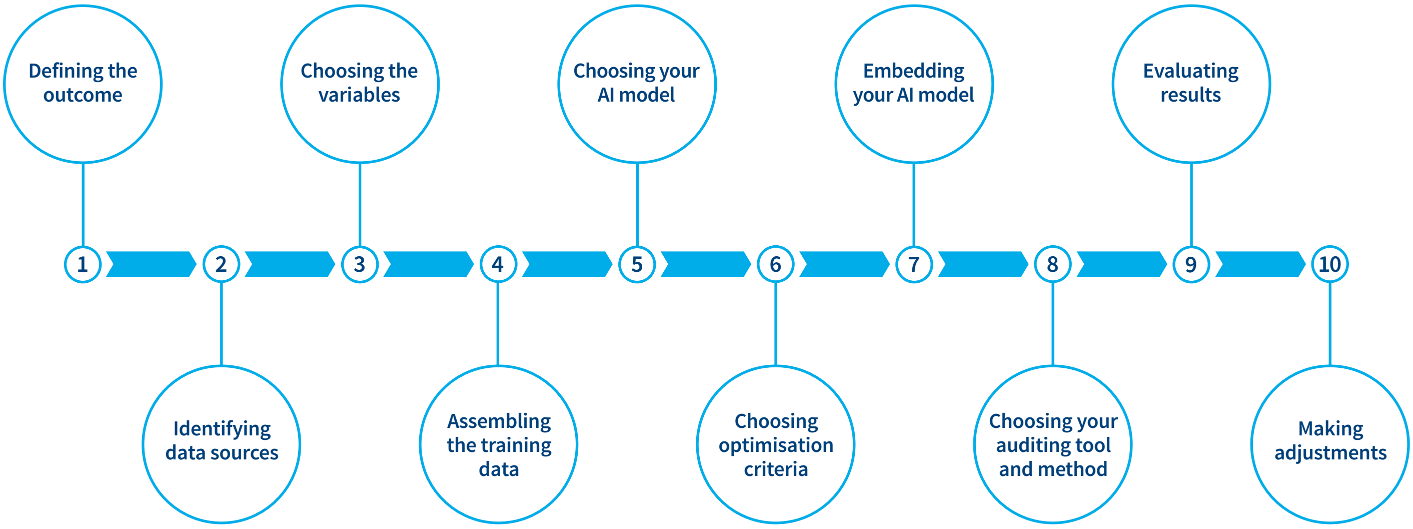
Best practice
Have you considered how insights from your technical audit or any other part of this assessment may enable you to contribute to advancing equality between the individuals and groups?

Best practice
Can you describe the positive steps you have taken to promote equality between the individuals and groups?

Best practice
Have you publicly disclosed, or made publicly available, a summary of key definitions, programming and training sources, and the methodologies of your AI system and any auditing tools you have used?

Best practice
Be prepared to answer questions about your decisions at each key decision-making stage in the flow chart below.

Key decision-making points



> Frameworks will shape whether AI undermines, or furthers, the pursuit of equality at work.

Institute for the Future of Work

Endnotes

1 The Royal Society, “AI and Work,” September 11, 2018,
https://royalsociety.org/topics-policy/projects/ai-and-work/;

FoW Commission, “The Future of Work Commission,” 2017,
https://d3n8a8pro7vhmx.cloudfront.net/campaigncountdown/pages/1052/attachments/original/1512946196/Future_of_Work_Commission_Report__December_2017.pdf?1512946196.

2 Jon Shields. “Over 98% of Fortune 500 Companies Use Applicant Tracking Systems (ATS),” *Jobscan Blog* (blog), June 20, 2018,
https://www.jobscan.co/blog/fortune-500-use-applicant-tracking-systems/;

Javier Sanchez-Monedero, Lina Dencik, and Lilian Edwards. “What Does It Mean to Solve the Problem of Discrimination in Hiring? Social, Technical and Legal Perspectives from the UK on Automated Hiring Systems”
(Conference on Fairness, Accountability, and Transparency (FAT* ’20), Barcelona, Spain, 2020),
http://arxiv.org/abs/1910.06144.

3 Polly Mosendz and Andrews Melin. “Bosses Panic-Buy Spy Software to Keep Tabs on Remote Workers,”
Bloomberg.Com, March 27, 2020,
https://www.bloomberg.com/news/features/2020-03-27/bosses-panic-buy-spy-software-to-keep-tabs-on-remote-workers.

4 Economist, ‘The coronavirus crisis thrusts corporate HR chiefs into the spotlight’ March 26 (2020). Available at:
https://www.economist.com/business/2020/03/26/the-coronavirus-crisis-thrusts-corporate-hr-chiefs-into-the-spotlight

5 Economist, ‘How covid-19 is driving public-sector innovation’
https://www.economist.com/britain/2020/04/03/how-covid-19-is-driving-public-sector-innovation (Accessed April 3, 2020).

6 Josh Simons. “Machine Learning at Work: Case Studies,” Institute for the Future of Work, February 26, 2020,
https://www.ifow.org/publications/2020/2/24/machine-learning-case-studies.
Forthcoming IFOW analysis of Retail and Transport sectors, supported by Trust for London.

7 Joshua Simons and Danielle Allen. “Harvard Ethics: COVID-19 Response White Papers,” 2020,
https://ethics.harvard.edu/covid-19-response.

8 Matt Craven et al. “Coronavirus’ Business Impact: Evolving Perspective,” McKinsey, March 30, 2020,
https://www.mckinsey.com/business-functions/risk/our-insights/covid-19-implications-for-business;

Andres Vinelli, Christian E. Weller, and Divya Vijay. “The Economic Impact of Coronavirus in the U.S. and Possible Economic Policy Responses,” Center for American Progress, March 6, 2020,
https://www.americanprogress.org/issues/economy/news/2020/03/06/481394/economic-impact-coronavirus-united-states-possible-economic-policy-responses/.

9 Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. “The allocation of talent and US economic growth.”
Econometrica 87, no. 5 (2019): 1439–1474.

10 Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen.
“Who becomes an inventor in America? The importance of exposure to innovation.”
The Quarterly Journal of Economics 134, no. 2 (2019): 647–713.

11 Fortune ‘25 ideas that will shape the 2020s’ December 19, 2019 9:30 AM GMT
https://fortune.com/longform/ideas-shape-2020s-tech-economy-markets-ai-health-work-society/
(Accessed April 16, 2020).

12 FAST are not the only principles, with many more recognised, for instance, in the OECD Principles for AI.
Miranda Bogen. “Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias” (Upturn, December 2018), 1,
https://www.upturn.org/reports/2018/hiring-algorithms/.

13 Andersson, Annika, Karin Hedström, and Elin Wihlborg.
“Automated Decision-Making and Legitimacy in Public Administration.”
In Scandinavian Workshop on Electronic Government (SWEG 2018), Copenhagen, Denmark, Jan. 31–Feb. 1, 2018. 2018.

Endnotes

14 ICO, “Guidance on the AI Auditing Framework,” Draft guidance for consultation (Information Commissioner’s Office, 2020), https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf;

ICO, “Big Data, Artificial Intelligence, Machine Learning and Data Protection” (Information Commissioner’s Office, 2017), https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf;

EC, “On Artificial Intelligence: A European Approach to Excellence and Trust,” White Paper (European Commission, February 19, 2020), https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf;

CE, “Addressing the Impacts of Algorithms on Human Rights: Draft Recommendation of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems” (Council of Europe, 2018), https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf;

Committee on Standards in Public Life, “Artificial Intelligence and Public Standards: Report,” February 2020, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF;

CDEI, “Review into Bias in Algorithmic Decision-Making,” Interim Report (Centre for Data Ethics and Innovation, 2018), https://www.gov.uk/government/publications/interim-reports-from-the-centre-for-data-ethics-and-innovation/interim-report-review-into-bias-in-algorithmic-decision-making.

15 Kaminski, Margot E., and Gianclaudio Malgieri. “Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations.” Available at SSRN 3456224 (2019). See also Recital 71 of GPDR – discrimination is mentioned in passing in the context of discussing accuracy and bias.

16 Centre for Data Ethics and Innovation, Landscape Summary: Bias in Algorithmic Decision Making. Availble at: https://www.gov.uk/government/publications/landscape-summaries-commissioned-by-the-centre-for-data-ethics-and-innovation.

17 Quote from p7 of Joshua Simons et al. “Equality through Transition,” Institute for the Future of Work, February 13, (2019) https://www.ifow.org/publications/2019/2/13/equality-through-transition.

18 Campolo, Alexander, and Kate Crawford. “Enchanted Determinism: Power without Responsibility in Artificial Intelligence.” Engaging Science, Technology, and Society 6 (2020): 1–19.

19 Virginia Eubanks. *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor* (New York, NY: St Martin’s Press, 2018);

Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016);

Joshua A. Kroll et al. “Accountable Algorithms,” *University of Pennsylvania Law Review* 165, no. 3 (2017): 633–705;

Frank Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2015);

Mike Ananny and Kate Crawford. “Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability,” *New Media & Society* 20, no. 3 (2018): 973–989.

20 Josh Simons. “Machine Learning at Work: Case Studies,” Institute for the Future of Work, 26 2020, https://www.ifow.org/publications/2020/2/24/machine-learning-case-studies;

Jordan Weissmann. “Amazon Created a Hiring Tool Using AI. It Immediately Started Discriminating Against Women.,” Slate Magazine, October 10, 2018, https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html;

Department of Housing and Urban Development, “Charge of Discrimination,” 2019, https://www.hud.gov/press/press_releases_media_advisories/HUD_No_19_035.

21 Reva B. Siegel. “Blind Justice: Why The Court Refused to Accept Statistical Evidence of Discriminatory Purpose in McCleskey v. Kemp - and Some Pathways for Change,” *Northwestern University Law Review* 112, no. 6 (2018): 1269–1291;

Barbara D. Underwood. “Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment,” *The Yale Law Journal* 88, no. 7 (1979): 1408–1448.

22 Within this body of work, the Turing Institute’s recent articulation of “Fairness, Accountability, Sustainability, and Transparency” (FAST) stand out. In this paper, we use this articulation as the best synthesis of the recent proliferation of principles around the governance of AI, including international principles.

23 Kaminski, Margot E., and Gianclaudio Malgieri. “Multi-layered explanations from algorithmic impact assessments in the GDPR.” In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 68–79. 2020.

Endnotes

24 David Leslie. “Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector” (The Alan Turing Institute, 2019), 11, https://doi.org/10.5281/zenodo.3240529;
ICO, “Guidance on the AI Auditing Framework,” 36.

25 Kate Crawford et al. “2019 Report” (New York: AI Now, 2019), 16, https://ainowinstitute.org/AI_Now_2019_Report.pdf;
Brent Mittelstadt, “Principles Alone Cannot Guarantee Ethical AI,” *ArXiv.Org*, 2020, https://doi.org/10.1038/s42256-019-0114-4.

26 Verma, Sahil, and Julia Rubin. “Fairness definitions explained.” In *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, pp. 1-7. IEEE, 2018.

27 Jack M. Balkin and Reva B. Siegel. “The American Civil Rights Tradition: Anticlassification or Antisubordination,” *Issues in Legal Scholarship* 2, no. 1 (2003);
Muhammad Bilal Zafar et al. “Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment,” October 26, 2016, http://arxiv.org/abs/1610.08452.

28 HUD, “Proposed Rule,” Pub. L. No. 84 FR 42854, Docket No. FR-6111-P-02 24 CFR 100 (2019), https://www.federalregister.gov/documents/2019/08/19/2019-17542/huds-implementation-of-the-fair-housing-acts-disparate-impact-standard#h-9.

29 Jon Kleinberg and Sendhil Mullainathan. “Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability,” September 12, 2018.

30 Sam Corbett-Davies and Sharad Goel. “The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning,” 2018, http://arxiv.org/abs/1808.00023;
Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. “Inherent Trade-Offs in the Fair Determination of Risk Scores,” *Proceedings of Innovations in Theoretical Computer Science (ITCS)*, 2017;
Jon Kleinberg and Sendhil Mullainathan. “Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability,” September 12, 2018;
Jon Kleinberg et al. “Discrimination in the Age of Algorithms,” 2019, http://arxiv.org/abs/1902.03731.

31 Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. “Inherent Trade-Offs in the Fair Determination of Risk Scores,” *Proceedings of Innovations in Theoretical Computer Science (ITCS)*, 2017;
Alexandra Chouldechova. “Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments,” *FATML 2016 Conference Paper*, 2016.

32 Kleinberg, Mullainathan, and Raghavan, Ibid.

33 Sorelle Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. “On the (Im)Possibility of Fairness,” *ArXiv.Org*, 2016, http://search.proquest.com/docview/2080432673/?pq-origsite=primo;
William Dietrich, Christina Mendoza, and Tim Brennan. “COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity” (Northpointe Inc., July 8, 2016), http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf;
Jiahao Chen et al. “Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved,” FAT ’19 (ACM, 2019), 339–348.

34 Robin Allen QC. Cloisters, 3 February 2020, “*Artificial intelligence, machine learning, Algorithms and discrimination law: The new frontier*” at Michael Rubenstein’s Annual Discrimination Law Conferences in Edinburgh and London. Available at: https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020.FINAL_-1.pdf (Accessed April 22, 2020).

35 Sheila Foster. “Causation in Antidiscrimination Law: Beyond Intent versus Impact,” Houston Law Review 41, no. 5 (2005): 1469–1548;
Richard W. Wright. “Causation in Tort Law.,” *California Law Review* 73 (1985): 1735–1956.

36 Matt J. Kusner et al. “Counterfactual Fairness,” March 20, 2017, http://arxiv.org/abs/1703.06856.

37 The Equality Act (2010) Explanatory Notes
http://www.legislation.gov.uk/ukpga/2010/15/notes/contents.
http://arxiv.org/abs/1910.06144.

Endnotes

[38] Frank Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2015);

[39] Albert (2019: 217–18) identifies 11 areas across the recruitment and selection process where AI-applications can be applied.

[40] For work on mitigation, see Friedler, Sorelle A., Carlos Scheidegger, Suresh Venkatasubramanian, Sonam Choudhary, Evan P. Hamilton, and Derek Roth. "A comparative study of fairness-enhancing interventions in machine learning." In Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 329–338. 2019.

[41] Robin Allen. "Articial Intelligence, Machine Learning, Algorithms and Discrimination Law: The New Frontier" (Discrimination Law in 2020, Congress House, 2020), https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020.FINAL_-1.pdf;

Javier Sanchez-Monedero, Lina Dencik, and Lilian Edwards. "What Does It Mean to Solve the Problem of Discrimination in Hiring? Social, Technical and Legal Perspectives from the UK on Automated Hiring Systems" (Conference on Fairness, Accountability, and Transparency (FAT* '20), Barcelona, Spain, 2020), http://arxiv.org/abs/1910.06144.

[42] Scholars have recently explored the extent and basis of underlying conceptual similarities between these approaches. It is important to be explicit about these tensions in the underlying idea of discrimination, as they are likely to become increasingly important to equality law and policy over the coming decades.

Hugh Collins and Tarunabh Khaitan. *Foundations of Indirect Discrimination Law* (Oxford: Hart Publishing, 2018);

Sophia Moreau and Deborah Hellman. *Philosophical Foundations of Discrimination Law* (Oxford: Oxford University Press, 2013).

[43] Solon Barocas and Andrew D. Selbst. "Big Data's Disparate Impact," *California Law Review* 104, no. 3 (June 1, 2016): 671–732;

Reva B. Siegel. "Blind Justice: Why The Court Refused to Accept Statistical Evidence of Discriminatory Purpose in McCleskey v. Kemp - and Some Pathways for Change," *Northwestern University Law Review* 112, no. 6 (2018): 1269–1291;

Balkin and Siegel. "The American Civil Rights Tradition"; Robin West, Civil Rights: *Rethinking Their Natural Foundation*, Cambridge Studies on Civil Rights and Civil Liberties (Cambridge: Cambridge University Press, 2019).

[44] Sam Corbett-Davies and Sharad Goel. "The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning," 2018, http://arxiv.org/abs/1808.00023.

[45] Ellen Sheng. Employee Privacy in The U.S. is at Stake as Corporate Surveillance Technology Monitors Workers' Every Move, CNBC (Apr. 15, 2019), https://www.cnbc.com/2019/04/15/employee-privacy-is-at-stake-as-surveillance-tech-monitors-workers.html cited in Nelson, Josephine, Management Culture and Surveillance (December 16, 2019). 43 Seattle U. L. Rev. 2, 631 (2020) (Berle XI symposium on Corporate Culture). Available at SSRN: https://ssrn.com/abstract=3504408

[46] Kaminski, Margot E., and Gianclaudio Malgieri. "Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations." (2019).

[47] This develops IFOW's recommendation that governments and businesses be required to conduct Equality Impact Assessments (EIAs) in Joshua Simons et al., "Equality through Transition," Institute for the Future of Work, February 13, 2019, 15, https://www.ifow.org/publications/2019/2/13/equality-through-transition. The final version of EIAs will be published in the final report of the Equality Task Force.

[48] Allen, Robin. "Artificial Intelligence, Machine Learning, Algorithms and Discrimination Law: The New Frontier." Congress House, 2020. https://482pe539799u3ynseg2hl1r3-wpengine.netdna-ssl.com/wp-content/uploads/2020/02/Discrimination-Law-in-2020.FINAL_-1.pdf.

[49] Committee on Standards in Public Life, "Artificial Intelligence and Public Standards: Report," 46–47.

[50] Committee on Standards in Public Life, 47.

[51] Council of Europe Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence MSI-AUT 'Addressing the impacts of Algorithms on Human Rights Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems' (2018). Available at: https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf (Accessed April 16, 2020).

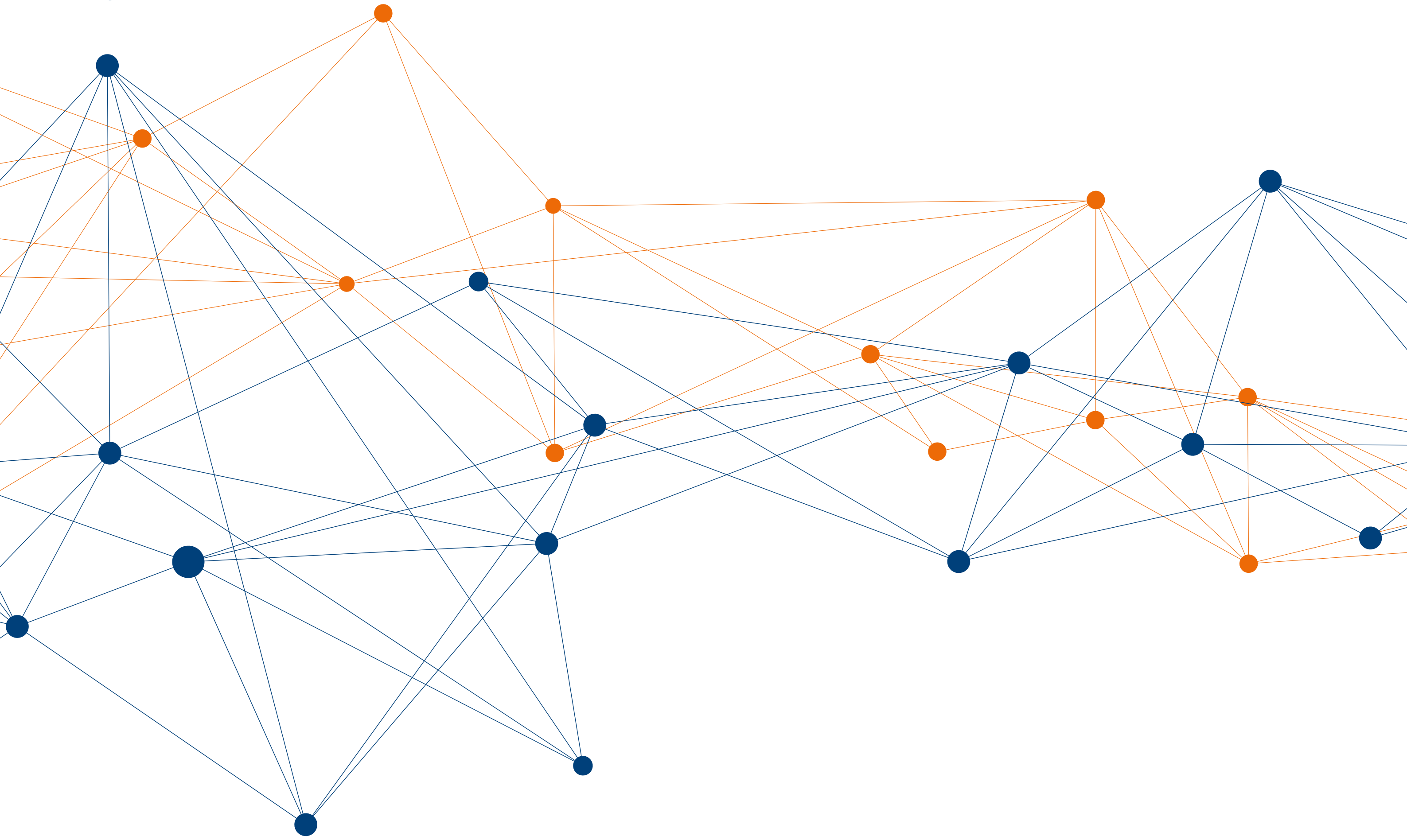
[52] Sanchez-Monedero, Javier, Lina Dencik, and Lilian Edwards. "What Does It Mean to Solve the Problem of Discrimination in Hiring? Social, Technical and Legal Perspectives from the UK on Automated Hiring Systems." Barcelona, Spain, 2020.

[53] Committee on Standards in Public Life, 47.

IFOW Institute *for the* Future of Work

Somerset House, Strand
London WC2R 1LA
T +44 (0)20 3701 7633

www.ifow.org
@_futureofwork



With thanks to Dr Lina Dencik, Oliver Buckley, Andrew Pakes, Robin Allen QC, Samuel Atwell, Paula Hagan, Jenny Brennan, and Fraser Watson.

Registered charity number 1180718