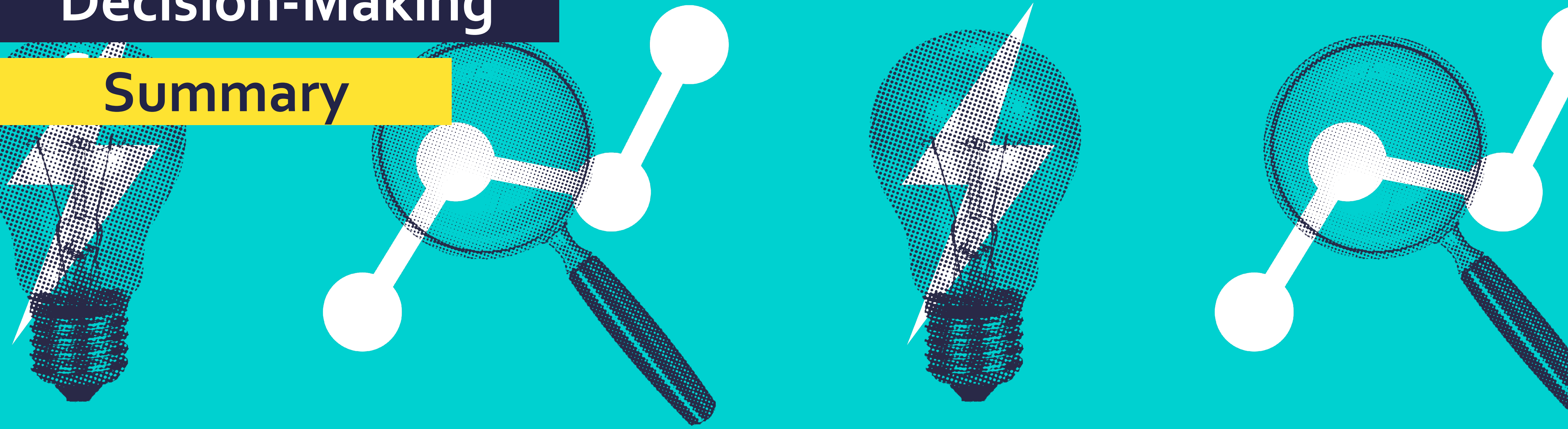
Centre for
Data Ethics
and Innovation

Review into Bias in Algorithmic Decision-Making

Summary



27th November 2020

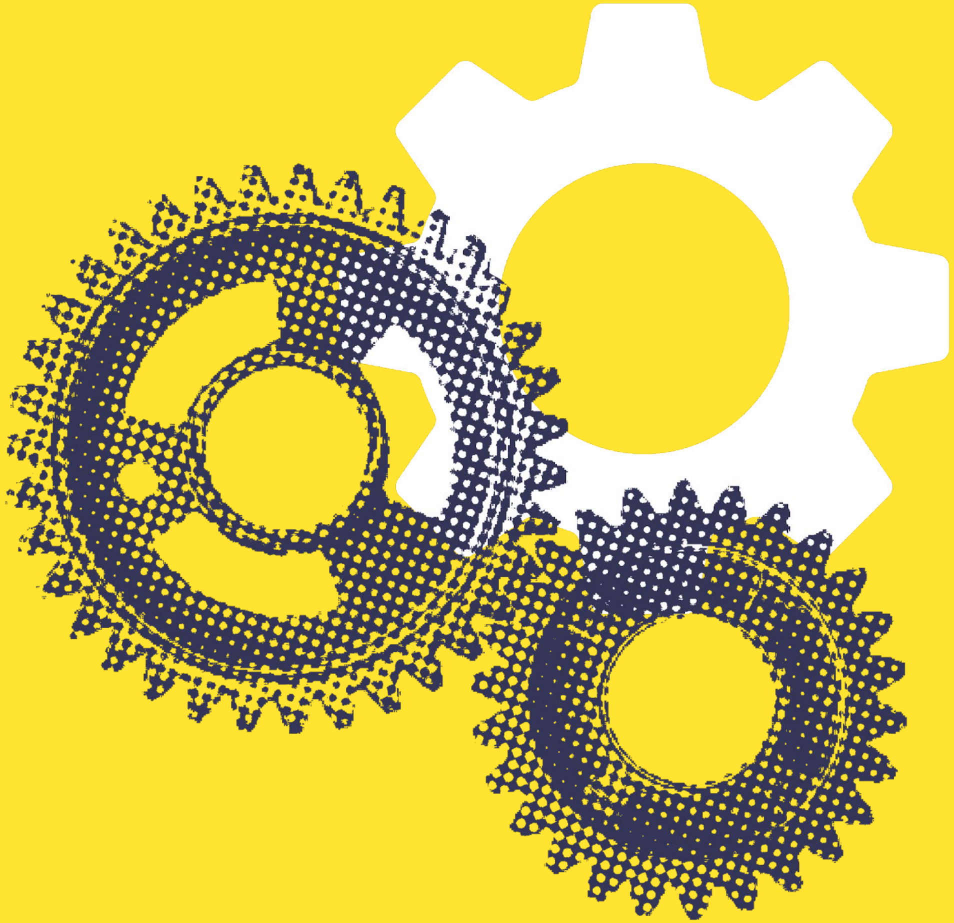
Centre for
Data Ethics
and Innovation

About the CDEI

The Centre for Data Ethics and Innovation (CDEI) is an independent expert advisory body, set up and tasked by the UK government to investigate and advise on how we maximise the benefits of data-driven technologies.

Our goal is to create the conditions needed for ethical innovation, development and deployment of data-driven technology, which the public can trust. We draw on expertise and perspectives from stakeholders across society to identify the risks and benefits posed by data-driven technologies. This enables us to provide advice and recommendations to regulators, industry and government to mitigate risk and incentivise ethical practices. The government has committed to respond publicly to our recommendations.

More information about the CDEI can be found at www.gov.uk/cdei



The report

Part I: Introduction

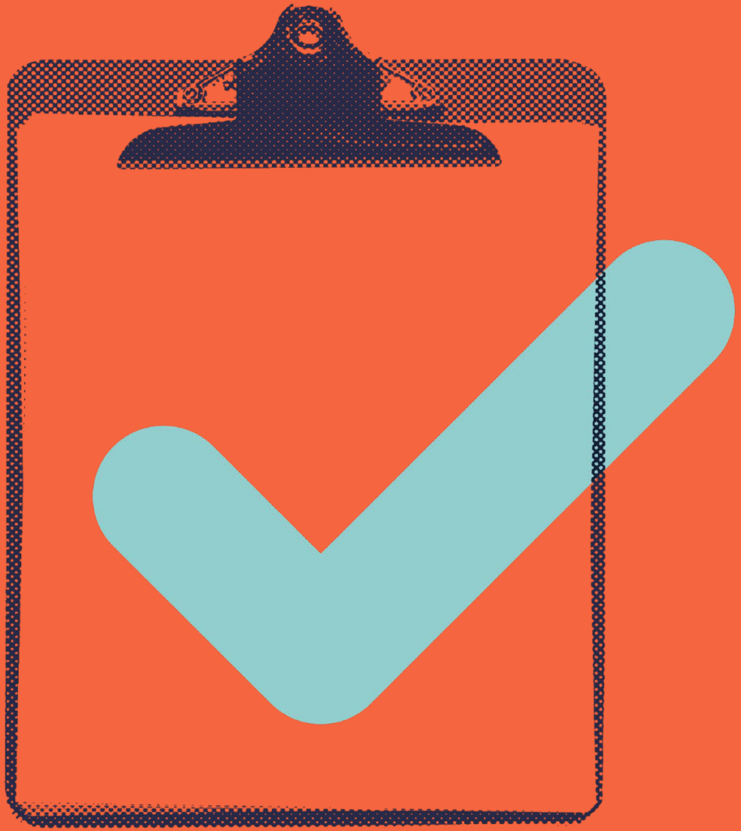
1. Background and scope
2. The issue

Part II: Sector reviews

3. Recruitment
4. Financial services
5. Policing
6. Local government

Part III: Addressing the challenges

7. Enabling fair innovation
8. The regulatory environment
9. Transparency in the public sector
10. Next steps & future challenges



This presentation summarises CDEI's final report. The full version of this report can be found on CDEI's website.

Part I: Introduction

In the October 2018 Budget, the Chancellor announced that we would investigate the potential bias in decisions made by algorithms. This review formed a key part of our 2019/2020 work programme, though completion was delayed by the onset of COVID-19. This is the final report of the CDEI's review and includes a set of formal recommendations to the government.

Government tasked us to draw on expertise and perspectives from stakeholders across society to provide recommendations on how they should address this issue. We also provide advice for regulators and industry, aiming to support responsible innovation and help build a strong, trustworthy system of governance. The government has committed to consider and respond publicly to our recommendations.

1. **Background and scope**
2. **The issue**

Background and approach

Scope and focus

This review focuses on decisions where potential bias seems to represent a significant and imminent ethical risk:

- Where algorithms have the potential to make or inform a significant decision that directly affects an individual human being
- Where algorithmic decision-making is being used now, or likely to be soon
- Where algorithmic decision making changes ethical risks
- Where decisions are potentially biased rather than other forms of unfairness such as arbitrariness or unreasonableness.

Sector Approach

We chose four initial areas of focus to illustrate the range of issues. These were **policing, financial services, recruitment and local government.** These sectors:

- Involve making **decisions at scale about individuals** which involve **significant potential impacts** on those individuals' lives.
- Have **growing interest in the use of algorithmic decision-making** tools, including machine learning.
- Have **evidence of historic bias** in decision-making within these sectors, leading to risks of this being perpetuated by the introduction of algorithms.

Cross-cutting themes

We set out three key cross-cutting questions in our interim report, which we have sought to address on a cross-sector basis:

- **Data**: Do organisations and regulators have access to the data they require to adequately identify and mitigate bias?
- **Tools and techniques**: What statistical and technical solutions are available now or will be required in future to identify and mitigate bias and which represent best practice?
- **Governance**: Who should be responsible for governing, auditing and assuring these algorithmic decision-making systems?

Evidence Base

This review draws on a range of evidence, including:

- A landscape summary of the academic and policy literature
- An open call for evidence
- Polling and a behavioural science experiment on public attitudes
- Broad engagement with industry, including a series of semi-structured interviews with finance & recruitment companies.
- Commissioned research on bias mitigation techniques and data analytics in policing

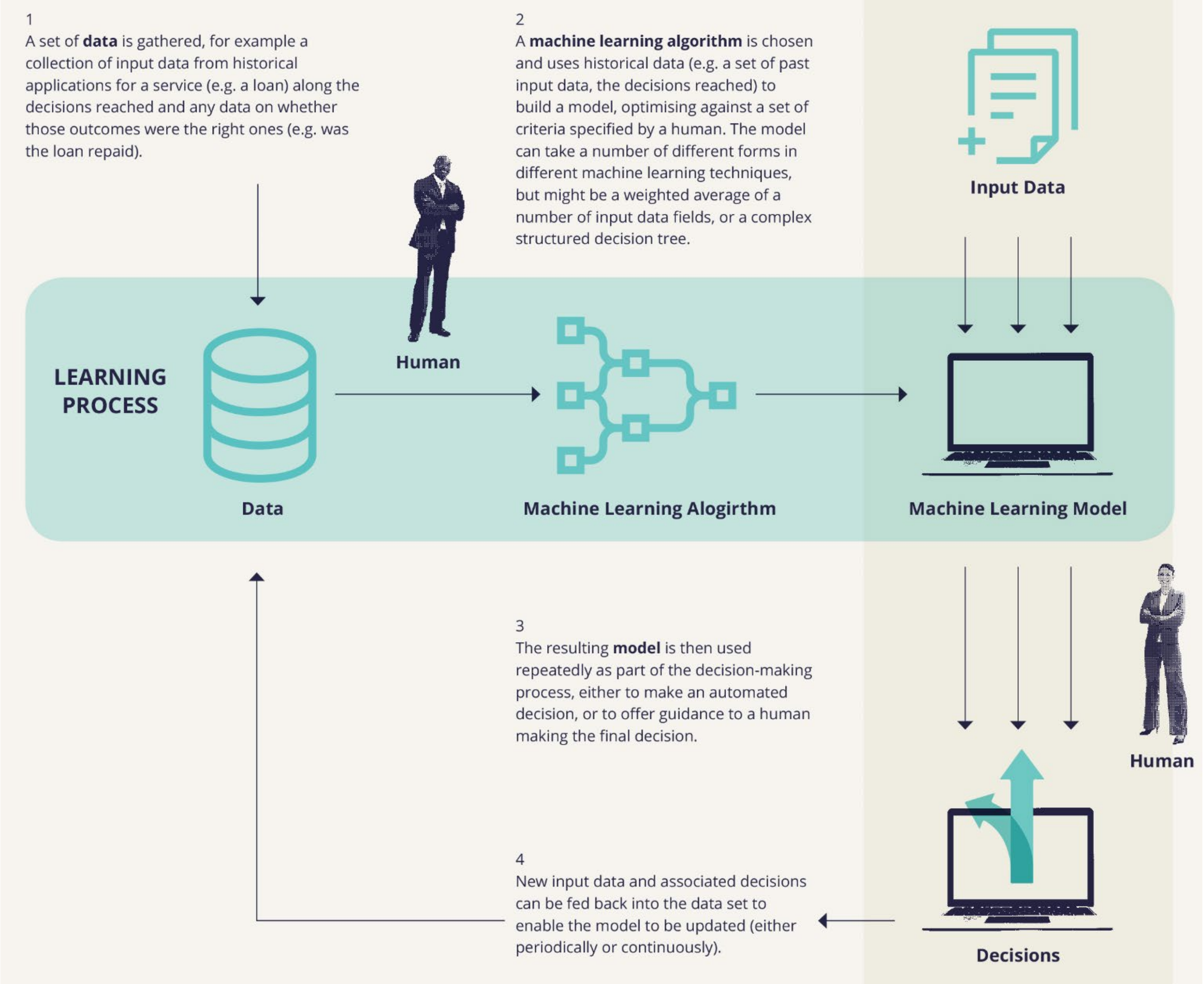
Using algorithms in decision-making

Algorithms are structured processes, which have long been used to aid human decision-making.

Developments in machine learning techniques and an exponential growth in the availability of data have allowed for more sophisticated algorithmic decisions.

Algorithms often do not represent the complete decision-making process: it is humans that decide on the objectives it is trying to meet, the data available to it, and how the output is used.

Organisations and their leaders are responsible for their decisions — whether they have been made by an algorithm or a team of humans.



1
A set of **data** is gathered, for example a collection of input data from historical applications for a service (e.g. a loan) along the decisions reached and any data on whether those outcomes were the right ones (e.g. was the loan repaid).

2
A **machine learning algorithm** is chosen and uses historical data (e.g. a set of past input data, the decisions reached) to build a model, optimising against a set of criteria specified by a human. The model can take a number of different forms in different machine learning techniques, but might be a weighted average of a number of input data fields, or a complex structured decision tree.

3
The resulting **model** is then used repeatedly as part of the decision-making process, either to make an automated decision, or to offer guidance to a human making the final decision.

4
New input data and associated decisions can be fed back into the data set to enable the model to be updated (either periodically or continuously).

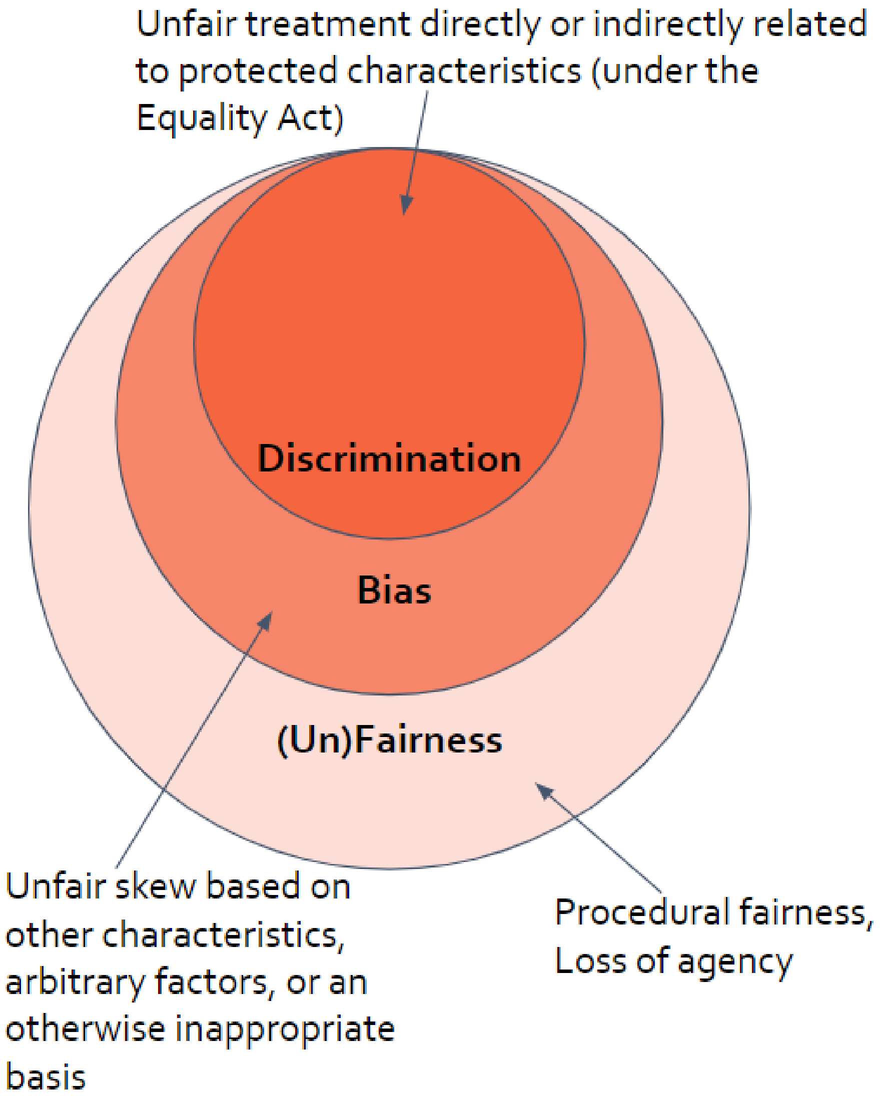
Bias, discrimination & fairness

The growth in algorithmic decision-making has been accompanied by significant concerns about **bias**; that the use of algorithms can cause a systematic skew in decision-making that results in unfair outcomes. There is clear evidence that algorithmic bias can occur, whether through entrenching previous human biases or introducing new ones.

Some forms of bias constitute **discrimination** under the Equality Act 2010, namely when bias leads to unfair treatment based on certain **protected characteristics**. There are also other kinds of algorithmic bias that are non-discriminatory, but still lead to unfair outcomes.

Fairness is about much more than the absence of bias; fair decisions need to also be non-arbitrary, reasonable, consider equality implications and respect the circumstances and personal agency of individuals.

There are multiple concepts of fairness, some of which are incompatible and can be ambiguous. In human decisions we can often accept ambiguity and allow human judgement to consider complex reasons for a decision. In contrast, algorithms are unambiguous.


Unfair treatment directly or indirectly related to protected characteristics (under the Equality Act)
Discrimination
Bias
(Un)Fairness
Unfair skew based on other characteristics, arbitrary factors, or an otherwise inappropriate basis
Procedural fairness, Loss of agency

Fairness in algorithmic decision-making

Fairness in decision-making

We cannot separate the question of algorithmic bias from the question of fair decision-making more broadly. It is important that the overall decision-making process is fair, not merely that algorithms are unbiased.

Data gives us a powerful weapon to address this. Good use of data can identify where bias is occurring, help us investigate why, and measure whether our efforts to combat it are effective.

However, data can also make things worse. New forms of data driven decision-making have surfaced numerous examples where algorithms have entrenched or amplified historic biases, or even created new forms of bias or unfairness. **This highlights the urgent need for the world to do better in using algorithms in the right way: to promote fairness, not undermine it.**

We now have the opportunity to adopt a more rigorous and proactive approach to identifying and mitigating bias in key areas of life. **Fairness through unawareness is often not enough to prevent bias:** ignoring protected characteristics is insufficient to prevent algorithmic bias and it can prevent organisations from identifying and addressing bias.

Why algorithms are different

There are plenty of fairness issues with human decision-making; **but** some of challenges with algorithms are different.

Organisations often rely on human decision-makers to interpret guidance appropriately and apply **human judgement** when required, especially in unusual cases. An algorithm can't do this: it will optimise against an objective without balance if told to do so.

Despite concerns about 'black box' algorithms, in some ways algorithms can be more transparent than human decision-making. Unlike human decision-making, with algorithms it's possible to test a system's response to changing inputs (even though the underlying logic can sometimes be opaque for both humans and machine learning algorithms!).

Algorithms are also consistent. This can be a good thing, but it is also a risk. While a single biased human can only make so many biased decisions; the impact of an algorithm can scale across a much larger number.

It is critical for successful innovation that algorithms are used in a way that is both fair, and seen by the public to be fair.

Part II: Sector reviews

The ethical questions in relation to bias in algorithmic decision-making vary depending on the context and sector. We therefore chose four initial areas of focus to illustrate the range of issues.

The sectors chosen have the following in common:

- **They involve making decisions at scale about individuals which involve significant potential impacts on those individuals' lives.**
- **There is a growing interest in the use of algorithmic decision-making tools in these sectors, including those involving machine learning in particular.**
- **There is evidence of historic bias in decision-making within these sectors, leading to risks of this being perpetuated by the introduction of algorithms.**

There are of course other sectors that we could have considered; these were chosen as a representative sample across the public and private sector, not because we have judged that the risk of bias is most acute in these specific cases.

3. Recruitment
4. Financial services
5. Policing
6. Local government

Recruitment

Current landscape

The use of algorithms in recruitment is becoming more widespread.

When developed responsibly, data-driven tools can improve processes and redress human bias. However, when using historical employment data, these tools may replicate or exacerbate structural inequality.

Analysis and gaps

Rigorous testing of new tools is necessary to ensure platforms do not unintentionally discriminate against groups of people. This will require collecting demographic data to test for bias.

We found that organisations are unsure of about how they should achieve their responsibilities across both the Equality Act 2010 and Data Protection Act 2018.

Recommendation 1

The **EHRC** should update their guidance on the application of the Equality Act 2010 to recruitment, to reflect issues associated with the use of algorithms, in collaboration with consumer and industry bodies.

Recommendation 2

The **ICO** should work with industry to understand why current guidance is not being consistently applied, and consider updates to guidance (e.g. in the Employment Practices Code), greater promotion of existing guidance, or other action as appropriate.

Advice to industry

Organisations should carry out Equality Impact Assessments to understand how their models perform for candidates with different protected characteristics, including intersectional analysis for those with multiple protected characteristics.

Future CDEI work

CDEI will consider how it can work with relevant organisations to assist with developing guidance on applying the Equality Act 2010 to algorithmic recruitment.

Financial services

Current landscape

Financial services organisations have long used data to support their decision-making. They range from being highly innovative to more risk averse in their use of new algorithmic approaches.

The regulatory picture is clearer in financial services than in the other sectors we have looked at. The **Financial Conduct Authority (FCA)** is the lead regulator and is conducting work to understand the impact and opportunities of innovative uses of data and AI in the sector.

Analysis and gaps

Explainability of models used in financial service is key, both to build trust with customers and to address potentially biased outcomes.

There are mixed views and approaches amongst financial organisations on the collection and use of protected characteristics data and this affects the ability of organisations to check for bias.

Future CDEI work

CDEI will be an observer on the Financial Conduct Authority and Bank of England’s AI Public Private Forum which will explore means to support the safe adoption of machine learning and artificial intelligence within financial services.

Policing (1/2)

Current landscape

Police adoption of algorithmic decision-making is at an early stage in the UK, with very few tools currently in operation in the UK. Levels of usage and approaches to managing ethical risks vary greatly across police forces. Police forces have identified opportunities to use data analytics and AI at scale to better allocate resources, but there is a significant risk that without sufficient care, systematically unfair outcomes could occur.

Analysis and gaps

The use of algorithms to support police decision-making introduces new issues around the balance between security, privacy and fairness. There is a clear need for strong democratic oversight and meaningful public engagement on the acceptable uses of police technology.

The responsibility for ethical use of data in policing is fragmented. No one body is fully empowered or resourced to take ownership. There is strong momentum at the national level around data ethics in policing, but clearer leadership will be necessary in the long term. A clearer steer is required from the **Home Office.**

Policing (2/2)

Recommendation 3

The **Home Office** should define clear roles and responsibilities for national policing bodies with regards to data analytics and ensure they have access to appropriate expertise and are empowered to set guidance and standards. As a first step, the Home Office should ensure that work underway by the National Police Chiefs' Council and other policing stakeholders to develop guidance and ensure ethical oversight of data analytics tools is appropriately supported.

Future CDEI work

CDEI will be applying and testing its draft ethics framework for police use of data analytics with police partners on real-life projects.

Advice to police forces and suppliers

- Police forces should conduct an integrated impact assessment before investing in new data analytics software as a full operational capability, to establish a clear legal basis and operational guidelines for use of the tool. Further details of what the integrated impact assessment should include are set out in the report we commissioned from RUSI.
- Police forces should classify the output of statistical algorithms as a form of police intelligence, alongside a confidence rating indicating the level of uncertainty associated with the prediction.
- Police forces should ensure that they have appropriate rights of access to algorithmic software and national regulators should be able to audit the underlying statistical models if needed (for instance, to assess risk of bias and error rates). Intellectual property rights must not be a restriction on this scrutiny.

Local government

Current landscape

Local authorities are increasingly using data to inform decision-making and target services more effectively. The use of data-driven tools is still at an early stage.

While these tools present considerable opportunities to improve services, they should not be considered a silver bullet for funding challenges. In some cases use of these tools will require significant additional investment to fulfil their potential and possible increase in demand for services.

Recommendation 4

Government should develop national guidance to support local authorities to legally and ethically procure or develop algorithmic decision-making tools in areas where significant decisions are made about individuals, and consider how compliance with this guidance should be monitored.

Analysis and gaps

Data infrastructure and data quality are significant barriers to developing and deploying data-driven tools; investing in these is necessary before developing more advanced systems.

National guidance is needed to support local authorities to develop and use data-driven tools ethically, with specific guidance addressing how to identify and mitigate biases. There is also a need for wider sharing of best practice between local authorities.

Future CDEI Work

CDEI is exploring how best to support local authorities to responsibly and ethically develop data-driven technologies, including possible partnerships with both central and local government.

Part III: Addressing the challenges

In Part I we surveyed the issue of bias in algorithmic decision-making, and in Part II we studied the current state in more detail across four sectors. Here, we move on to identify how some of the challenges we identified can be addressed, the progress made so far, and what needs to happen next.

There are three main areas to consider:

- The **enablers** needed by organisations building and deploying algorithmic decision-making tools to help them do this in a fair way.
- The **regulatory** levers, both formal and informal, needed to incentivise organisations to do this, and create a level playing field for ethical innovation.
- How the **public sector**, as a major developer and user of data-driven technology, can show leadership in this area through transparency.

There are inherent links between these areas. Creating the right incentives can only succeed if the right enablers are in place to help organisations act fairly, but conversely there is little incentive for organisations to invest in tools and approaches for fair decision-making if there is insufficient clarity on the expected norms.

7. **Enabling fair development**
8. **The regulatory environment**
9. **Transparency in the public sector**
10. **Next steps & future challenges**

Addressing algorithmic bias (1/2)

Lots of good work is happening to try to make decision-making fair, but there remains a long way to go. We see the status quo as follows:

Opportunities	Challenges
Impact on bias: Algorithms could help to address bias.	Building algorithms that replicate existing biased mechanisms will embed or even exacerbate existing inequalities.
Measurement of bias: More data available than ever before to help organisations understand the impacts of decision-making.	Collection of protected characteristic data is very patchy, with significant perceived uncertainty about ethics, legality, and the willingness of individuals to provide data. There are uncertainties concerning the legality and ethics of inferring protected characteristics . Most decision processes (whether using algorithms or not) exhibit bias in some form and will fail certain tests of fairness. The law offers limited guidance to organisations on adequate ways to address this.
Mitigating bias: Lots of academic study and open source tooling available to support bias mitigation.	Relatively limited understanding of how to use these tools in practice to support fair, end-to-end, decision-making. A US-centric ecosystem where many tools do not align with UK equality law. Uncertainty about usage of tools, and issues on legality of some approaches under UK law. Perceived trade-offs with accuracy (though often this may suggest an incomplete notion of accuracy).
Expert support: A range of consultancy services are available to help with these issues.	An immature ecosystem, with no clear industry norms around these services, the relevant professional skills, or important legal clarity.

Addressing algorithmic bias (2/2)

Opportunities	Challenges
Workforce diversity: Strong stated commitment from government and industry to improving diversity	Still far too little diversity in the tech sector.
Leadership & governance: Many organisations understand the strategic drivers to act fairly and proactively in complying with data protection obligations	Recent focus on data protection (due to the arrival of GDPR), and especially privacy and security aspects of this, risks de-prioritisation of fairness and equality issues (even though these are also required in GDPR). Identifying historical or current bias in decision-making is not a comfortable thing for organisations to do. There is a risk that public opinion will penalise those who proactively identify and address bias.
Transparency: Transparency about the use and impact of algorithmic decision-making would help to drive greater consistency	There are insufficient incentives for organisations to be more transparent and risks to going alone. There is a danger of creating requirements that create public perception risks for organisations even if they would help reduce risks of biased decision-making . The UK public sector has identified this issue, but could do more to lead through its own development and use of algorithmic decision-making
Regulation: Good regulation can support ethical innovation	Not all regulators are currently equipped to deal with the challenges posed by algorithms. There is continued nervousness in industry around the implications of GDPR. The ICO has worked hard to address this, and recent guidance will help, but there remains a way to go to build confidence on how to interpret GDPR in this context.

Enabling fair innovation: findings

The challenge at hand

Many organisations are unsure how to address bias in practice. Support is needed to help them **consider, measure, and mitigate unfairness**.

Data is needed to monitor outcomes and identify bias, but **data on protected characteristics is not available often enough**. One cause is an incorrect belief that data protection law prevents collection or usage of this data.

There are some other genuine challenges in collecting this data, and more innovative thinking is needed in this area - such as trusted third party intermediaries.

More effective governance is needed around algorithmic decision making tools.

Approaches and gaps

Improving diversity in technology development is an important part of protecting against certain forms of bias. **Government and industry efforts to improve this must continue, and need to show results.**

The machine learning community has developed multiple techniques to measure and mitigate algorithmic bias. **Organisations should be encouraged to deploy methods that help to address algorithmic bias and discrimination.**

However, there is little guidance on how to choose the right methods, or how to embed them into development and operational processes.

Bias mitigation cannot be treated as a purely technical issue. It requires careful consideration of the wider policy, operational and legal context. There is insufficient legal clarity concerning novel techniques in this area, and some may not be compatible with equality law.

How should organisations address algorithmic bias?

Guidance to organisation leaders and boards

Those responsible for governance of organisations deploying or using algorithmic decision-making tools to support significant decisions about individuals should ensure that leaders are in place with accountability for:

- Understanding the capabilities and limits of those tools
- Considering carefully whether individuals will be fairly treated by the decision-making process that the tool forms part of
- Making a conscious decision on appropriate levels of human involvement in the decision-making process
- Putting structures in place to gather data and monitor outcomes for fairness
- Understanding their legal obligations and having carried out appropriate impact assessments

This especially applies in the public sector when citizens often do not have a choice about whether to use a service, and decisions made about individuals can often be life-affecting.

Achieving this in practice will include:

Building internal capacity

- Developing appropriate skills and tools to identify and address bias (which is a multidisciplinary task, not only a technical one)
- Improving workforce diversity

Understand risks of bias

- Gathering and analysing data to measure potential bias
- Using appropriate bias mitigation techniques
- Engaging with affected stakeholders

Creating organisational accountability

- Setting clear accountability and ownership over the decisions made about ensuring fair decisions
- Providing transparency about where algorithms are used, and how those decisions are made
- Engaging with regulators and industry bodies to set standards and norms

Centre for
Data Ethics
and Innovation

Enabling fair innovation: recommendations

To Government

Recommendation 5

Government should continue to support and invest in programmes that facilitate greater diversity within the technology sector, building on its current programmes and developing new initiatives where there are gaps.

Recommendation 6

Government should work with **relevant regulators** to provide clear guidance on the collection and use of protected characteristic data in outcome monitoring and decision-making processes. They should then encourage the use of that guidance and data to address current and historic bias in key sectors.

Recommendation 7

Government and the Office of National Statistics should open the Secure Research Service more broadly, to a wider variety of organisations, for use in evaluation of bias and inequality across a greater range of activities.

Recommendation 8

Government should support the creation and development of data-focused public and private partnerships, especially those focused on the identification and reduction of biases and issues specific to under-represented groups. **The Office of National Statistics** and **Government Statistical Service** should work with these partnerships and regulators to promote harmonised principles of data collection and use into the private sector, via shared data and standards development.

To Regulators and Industry Bodies

Recommendation 9

Sector regulators and industry bodies should help create oversight and technical guidance for responsible bias detection and mitigation in their individual sectors, adding context-specific detail to the existing cross-cutting guidance on data protection, and any new cross-cutting guidance on the Equality Act.

Enabling fair innovation: advice & guidance

Advice to industry

Organisations building and deploying algorithmic decision-making tools should make increased diversity in their workforce a priority. This applies not just to data science roles, but also to wider operational, management and oversight roles. Proactive gathering and use of data in the industry is required to identify and challenge barriers for increased diversity in recruitment and progression, including into senior leadership roles.

Where organisations operating within the UK deploy bias detection or mitigation tools developed in the US, they must be mindful that **relevant equality law (along with that across much of Europe) is different.**

Where organisations face historical issues, attract significant societal concern, or otherwise believe bias is a risk, they will need to measure outcomes by relevant protected characteristics to detect biases in their decision-making, algorithmic or otherwise. They must then address any uncovered direct discrimination, indirect discrimination, or outcome differences by protected characteristics that lack objective justification.

In doing so, organisations should ensure that their mitigation efforts do not produce new forms of bias or discrimination. Many bias mitigation techniques, especially those focused on representation and inclusion, can legitimately and lawfully address algorithmic bias when used responsibly. However, some risk introducing positive discrimination, which is illegal under the Equality Act. Organisations should consider the legal implications of their mitigation tools, drawing on industry guidance and legal advice.

The regulatory environment

The challenge at hand

AI presents genuinely new challenges for regulation, and brings into question whether existing legislation and regulatory approaches can address these challenges sufficiently well. There is currently little case law or statutory guidance directly addressing discrimination in algorithmic decision-making.

The role of regulation

Regulation can help to address algorithmic bias by setting minimum standards, providing clear guidance that supports organisations to meet their obligations, and enforcement to ensure minimum standards are met.

The current regulatory landscape for algorithmic decision-making consists of the Equality & Human Rights Commission, the Information Commissioner's Office and sector regulators.

Analysis and gaps

At this stage, we do not believe that there is a need for a new specialised regulator or primary legislation to address algorithmic bias.

However, algorithmic bias means that the overlap between discrimination law, data protection law and sector regulations is increasingly important, especially given 1) the use of protected characteristics (PCs) data to measure and mitigate algorithmic bias; 2) the lawful use of bias mitigation techniques; 3) identifying new forms of bias beyond existing PCs; and 4) for sector-specific measures of algorithmic fairness beyond discrimination.

Existing regulators need to adapt their enforcement to algorithmic decision-making, and provide guidance on how regulated bodies can maintain compliance in an algorithmic age. Some regulators require new capabilities to enable them to respond to the challenges of algorithmic decision-making. While larger regulators with a greater digital remit may be able to grow these capabilities in-house, others will need external support.

The regulatory framework

Algorithmic decision making sits within an existing regulatory context, which provides the foundations to address algorithmic bias.

I. Cross-cutting principles of fairness set out in equality and data protection law .

EHRC (& other regulators under their Public Sector Equality Duty):
Equality Act 2006/2010, Human Rights Act 1998

ICO:
Data Protection Act 2018, GDPR

These laws set out existing rights and obligations but need translation to algorithmic decision-making.

Directly applicable to algorithmic decision-making, but relatively new and not widely understood in an algorithmic context.

II. Sector regulators establish and enforce standards of fairness in particular sectors.

Private Sector	Public Sector
FCA: *Principles for fair treatment*	**HMICFRS**
CMA: *Consumer Rights Act*	**Ofsted**
Others...	Others...

Some regulators and sectors have a long history and capability in addressing algorithmic decision-making, others will need to focus on this area as the use of algorithmic decision-making grows increasingly important.

III. Anticipatory governance where organisations need to reflect on, and meet, what fairness means in their specific context.

Tools:
Industry codes of practice
Impact assessments
Certification and accreditation
Customer engagement
Stakeholder engagement

There are tools that will help to establish and embed common standards. Organisations and accountable leaders will still need to make choices about what concepts of fairness apply, and reasonable efforts to make fair decisions.

The regulatory environment: recommendations

To Government

Recommendation 10

Government should issue guidance that clarifies the Equality Act responsibilities of organisations using algorithmic decision-making. This should include guidance on the collection of protected characteristics data to measure bias and the lawfulness of bias mitigation techniques.

Recommendation 11

Through the development of this guidance and its implementation, **government** should assess whether it provides both sufficient clarity for organisations on meeting their obligations, and leaves sufficient scope for organisations to take actions to mitigate algorithmic bias. If not, **government** should consider new regulations or amendments to the Equality Act to address this.

Recommendation 12

The **EHRC** should ensure that it has the capacity and capability to investigate algorithmic discrimination. This may include EHRC reprioritising resources to this area, EHRC supporting other regulators to address algorithmic discrimination in their sector, and additional technical support to the EHRC.

Recommendation 13

Regulators should consider algorithmic discrimination in their supervision and enforcement activities, as part of their responsibilities under the Public Sector Equality Duty.

The regulatory environment: recommendations & advice

Centre for
Data Ethics
and Innovation

Recommendations to Regulators & Industry Bodies

Recommendation 14

Regulators should develop compliance and enforcement tools to address algorithmic bias, such as impact assessments, audit standards, certification and/or regulatory sandboxes.

Recommendation 15

Regulators should coordinate their compliance and enforcement efforts to address algorithmic bias, aligning standards and tools where possible. This could include jointly issued guidance, collaboration in regulatory sandboxes, and joint investigations.

Advice to Industry

Industry bodies and standards organisations should develop the ecosystem of tools and services to enable organisations to address algorithmic bias, including sector specific standards, auditing and certification services for both algorithmic systems and the organisations and developers who create them.

Future CDEI Work

The CDEI plans to grow its ability to provide expert advice and support to regulators, in line with our existing terms of reference. This will include supporting regulators to coordinate efforts to address algorithmic bias and to share best practice. The CDEI will monitor the development of algorithmic decision-making and the extent to which new forms of discrimination or bias emerge. This will include referring issues to relevant regulators, and working with government if issues are not covered by existing regulations.

Public sector transparency

The challenge at hand

Making decisions about individuals is a core responsibility of many parts of the public sector, and there is increasing recognition of the opportunities offered through the use of data and algorithms in decision-making.

The use of technology should never reduce real or perceived accountability of public institutions to citizens. In fact, it offers opportunities to improve accountability and transparency, especially where algorithms have significant effects on significant decisions about individuals.

Approaches and gaps

A range of transparency measures already exist around current decision-making processes. Given the current limited level adoption of algorithmic decision-making in the public sector, there is a window of opportunity to ensure that we get transparency right as adoption increases.

The supply chain that delivers an algorithmic decision-making tool will often include one or more suppliers external to the public body ultimately responsible for the decision-making itself. While the ultimate accountability for fair decision-making always sits with the public body, there is limited maturity or consistency in contractual mechanisms to place responsibilities in the right place in the supply chain.

Public sector transparency: recommendations

Recommendation 16

Government should place a mandatory transparency obligation on all public sector organisations using algorithms that have a significant influence on significant decisions affecting individuals. Government should conduct a project to scope this obligation more precisely, and to pilot an approach to implement it, but it should require the proactive publication of information on how the decision to use an algorithm was made, the type of algorithm, how it is used in the overall decision-making process, and steps taken to ensure fair treatment of individuals.

Recommendation 17

Cabinet Office and **the Crown Commercial Service** should update model contracts and framework agreements for public sector procurement to incorporate a set of minimum standards around ethical use of AI, with particular focus on expected levels of transparency and explainability, and ongoing testing for fairness.

Advice to industry

Industry should follow existing public sector guidance on transparency, principally within the Understanding AI Ethics and Safety guidance developed by the Office for AI, the Alan Turing Institute and the Government Digital Service, which sets out a process-based governance framework for responsible AI innovation projects in the UK public sector.

Future CDEI work

CDEI will support the **Government Digital Service** as they seek to scope and pilot an approach to transparency.

Next steps & future challenges

Centre for Data Ethics and Innovation

Framing our work

Recognising the breadth of this complex and evolving field, this report has focused heavily on surveying the maturity of the landscape, identifying the gaps, and setting out some concrete next steps.

Some of the next steps fall within CDEI's remit, and we are keen to help industry, regulators and government in taking forward the practical delivery work to address the issues we have identified and future challenges which may arise.

Government, industry bodies and regulators need to give more help to organisations building and deploying algorithmic decision-making tools on how to interpret the Equality Act in this context.
Drawing on the understanding built up through this review, CDEI is happy to support several aspects of the work in this space by, for example:

Potential next steps

Supporting the development of guidance

- Supporting the development of any guidance on the application of the Equality Act to algorithmic decision-making.
- Supporting government on developing guidance on collection and use of protected characteristics to meet responsibilities under the Equality Act, and in identifying any potential future need for a change in the law, with an intent to reduce barriers to innovation.
- Drawing on the draft technical standards work produced in the course of this review and other inputs to **help industry bodies, sector regulators and government departments in defining norms for bias detection and mitigation**.

Supporting the public sector

- Supporting the **Government Digital Service** as they seek to scope and pilot an approach to transparency.
- Through the course of the review, a number of public sector organisations have expressed interest in working further with us to apply the general lessons learnt in specific projects. For example, we will be supporting a police force and a local authority as they develop practical governance structures to support responsible and trustworthy data innovation.

Next steps & future challenges

Potential next steps

Providing expertise

Growing our ability to provide expert advice and support to regulators, in line with our terms of reference, including supporting regulators to coordinate efforts to address algorithmic bias and to share best practice. As an example, we have been invited to take an observer role on the Financial Conduct Authority and Bank of England's AI Public Private Forum which will explore means to support the safe adoption of machine learning and AI within financial services, with an intent to both support that work, and draw lessons from a relatively mature sector to share with others.

Convening and coordinating

We have noted the need for an ecosystem of skilled professionals and expert supporting services to help organisations in getting fairness right, and provide assurance. Some of the development needs to happen organically, but we believe that action may be needed to catalyse this.

CDEI plans to bring together a diverse range of organisations with interest in this area, and identifying what would be needed to foster and develop a strong AI accountability ecosystem in the UK. This is both an opportunity to manage ethical risks for AI in the UK, but also to support innovation in an area where there is potential for UK companies to offer audit services worldwide.

Leadership

Looking across the work listed above, and the future challenges that will undoubtedly arise, **we see a key need for national leadership and coordination to ensure continued focus and pace in addressing these challenges across sectors**.

Concluding this review

Final reflections

Government should be clear on where it wants this coordination to sit. There are a number of possible locations; for example in central government directly, in a regulator or in CDEI. **Government should be clear on where responsibilities sit for tracking progress across sectors in this area, and driving the pace of change.** As CDEI agrees our future priorities with government, we hope to be able to support them in this area.

This review has been, by necessity, a partial look at a very wide field. Indeed, some of the most prominent concerns around algorithmic bias to have emerged in recent months have unfortunately been outside of our core scope, including facial recognition and the impact of bias within how platforms target content (considered in CDEI’s review into online targeting).

Our AI Monitoring function will continue to monitor the development of algorithmic decision-making and the extent to which new forms of discrimination or bias emerge. This will include referring issues to relevant regulators, and working with government if issues are not covered by existing regulations.

Experience from this review suggests that many of the steps needed to address the risk of bias overlap with those for tackling other ethical challenges, for example structures for good governance, appropriate data sharing, and explainability of models. We anticipate that we will return to issues of bias, fairness and equality through much of our future work, though likely as one cross-cutting ethical issue in wider projects.

Centre for
Data Ethics
and Innovation

Thank you for reading this summary

Please email comments to:

bias@cdei.gov.uk