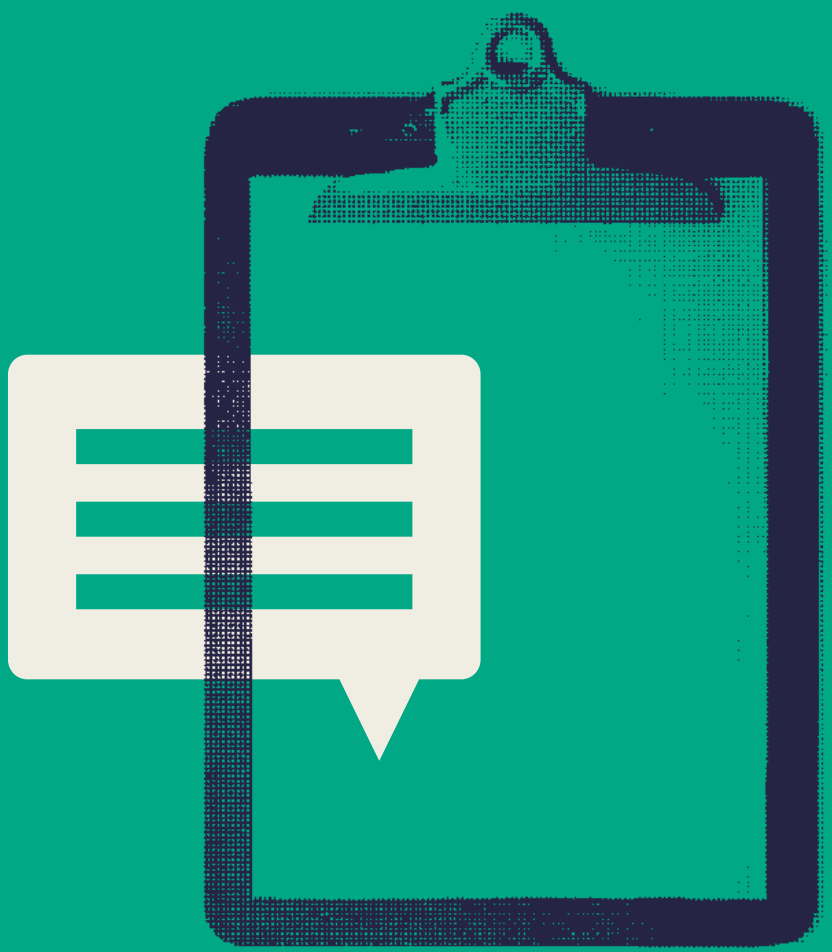
Executive summary



Unfair biases, whether conscious or unconscious, can be a problem in many decision-making processes. This review considers the impact that an increasing use of algorithmic tools is having on bias in decision-making, the steps that are required to manage risks, and the opportunities that better use of data offers to enhance fairness. We have focused on the use of algorithms in significant decisions about individuals, looking across four sectors (recruitment, financial services, policing and local government), and making cross-cutting recommendations that aim to help build the right systems so that algorithms improve, rather than worsen, decision-making.

It is well established that there is a risk that algorithmic systems can lead to biased decisions, with perhaps the largest underlying cause being the encoding of existing human biases into algorithmic systems. But the evidence is far less clear on whether algorithmic decision-making tools carry more or less risk of bias than previous human decision-making processes. Indeed, there are reasons to think that better use of data can have a role in making decisions fairer, if done with appropriate care.

When changing processes that make life-affecting decisions about individuals we should always proceed with caution. **It is important to recognise that algorithms cannot do everything.** There are some aspects of decision-making where human judgement, including the ability to be sensitive and flexible to the unique circumstances of an individual, will remain crucial.

Using data and algorithms in innovative ways can enable organisations to understand inequalities and to reduce bias in some aspects of decision-making. But there are also circumstances where using algorithms to make life-affecting decisions can be seen as unfair by failing to consider an individual's circumstances, or depriving them of personal agency. We do not directly focus on this kind of unfairness in this report, but note that this argument can also apply to human decision-making, if the individual who is subject to the decision does not have a role in contributing to the decision. History to date in the design and deployment of algorithmic tools has not been good enough. There are numerous examples worldwide of the introduction of algorithms persisting or amplifying historical biases, or introducing new ones. We must and can do better. Making fair and unbiased decisions is not only good for the individuals involved, but it is good for business and society. **Successful and sustainable innovation is dependent on building and maintaining public trust.** Polling undertaken for this review suggested that, prior to August's controversy over exam results, 57% of people were aware of algorithmic systems being used to support decisions about them, with only 19% of those disagreeing in principle with the suggestion of a "fair and accurate" algorithm helping to make decisions about them. By October, we found that awareness had risen slightly (to 62%), as had disagreement in principle (to 23%). This doesn't suggest a step change in public attitudes, but there is clearly still a long way to go to build **trust** in algorithmic systems. The obvious starting point for this is to ensure that algorithms are **trustworthy**.

The use of algorithms in decision-making is a complex area, with widely varying approaches and levels of maturity across different organisations and sectors. Ultimately, many of the steps needed to challenge bias will be context-specific. But from our work, we have identified a number of concrete steps for industry, regulators and government to take that can support ethical innovation across a wide range of use cases. **This report is not a guidance manual, but considers what guidance, support, regulation and incentives are needed to create the right conditions for fair innovation to flourish.**

It is crucial to take a broad view of the whole decision-making process when considering the different ways bias can enter a system and how this might impact on fairness. **The issue is not simply whether an algorithm is biased, but whether the overall decision-making processes are biased.** Looking at algorithms in isolation cannot fully address this.

It is important to consider bias in algorithmic decision-making in the context of all decision-making systems. Even in human decision-making, there are differing views about what is and isn't fair. But society has developed a range of standards and common practices for how to manage these issues, and legal frameworks to support this. Organisations have a level of understanding on what constitutes an appropriate level of due care for fairness. The challenge is to make sure that we can translate this understanding across to the algorithmic world, and apply a consistent bar of fairness whether decisions are made by humans, algorithms or a combination of the two. **We must ensure decisions can be scrutinised, explained and challenged so that our current laws and frameworks do not lose effectiveness, and indeed can be made more effective over time.** Significant growth is happening both in data availability and use of algorithmic decision-making across many sectors; **we have a window of opportunity to get this right and ensure that these changes serve to promote equality not to entrench existing biases.**

Sector reviews

The four sectors studied in Part II of this report are at different maturity levels in their use of algorithmic decision-making. Some of the issues they face are sector-specific, but we found common challenges that span these sectors and beyond.

In **recruitment** we saw a sector that is experiencing rapid growth in the use of algorithmic tools at all stages of the recruitment process, but also one that is relatively mature in collecting data to monitor outcomes. Human bias in traditional recruitment is well evidenced and therefore there is potential for data-driven tools to improve matters by standardising processes and using data to inform areas of discretion where human biases can creep in.

However, we also found that a clear and consistent understanding of how to do this well is lacking, leading to a risk that algorithmic technologies will entrench these inequalities. More guidance is needed on how to ensure that these tools do not unintentionally discriminate against groups of people, particularly when trained on historic or current employment data. Organisations must be particularly mindful to ensure they are meeting the appropriate legislative responsibilities around automated decision-making and reasonable adjustments for candidates with disabilities.

The innovation in this space has real potential for making **recruitment** fairer. However, given the potential risks, further scrutiny of how these tools work, how they are used and the impact they have on different groups, is required, along with higher and clearer standards of good governance to ensure that ethical and legal risks are anticipated and managed.

In **financial services,** we saw a much more mature sector that has long used data to support decision-making. Finance relies on making accurate predictions about peoples' behaviours, for example how likely they are to repay debts. However, specific groups are historically underrepresented in the financial system, and there is a risk that these historic biases could be entrenched further through algorithmic systems.

We found financial service organisations ranged from being highly innovative to more risk averse in their use of new algorithmic approaches. They are keen to test their systems for bias, but there are mixed views and approaches regarding how this should be done. This was particularly evident around the collection and use of protected characteristic data, and therefore organisations' ability to monitor outcomes.

Our main focus within financial services was on credit scoring decisions made about individuals by traditional banks. Our work found the key obstacles to further innovation in the sector included data availability, quality and how to source data ethically, available techniques with sufficient explainability, risk averse culture, in some parts, given the impacts of the financial crisis and difficulty in gauging consumer and wider public acceptance.

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

The use of data from non-traditional sources could enable population groups who have historically found it difficult to access credit, due to lower availability of data about them from traditional sources, to gain better access in future. At the same time, more data and more complex algorithms could increase the potential for the introduction of indirect bias via proxy as well as the ability to detect and mitigate it.

Adoption of algorithmic decision-making in the public sector is generally at an early stage. In **policing,** we found very few tools currently in operation in the UK, with a varied picture across different police forces, both on usage and approaches to managing ethical risks.

There have been notable government reviews into the issue of bias in policing, which is important context when considering the risks and opportunities around the use of technology in this sector. Again, we found potential for algorithms to support decision-making, but this introduces new issues around the balance between security, privacy and fairness, and there is a clear requirement for strong democratic oversight.

Police forces have access to more digital material than ever before, and are expected to use this data to identify connections and manage future risks. The £63.7 million funding for police technology programmes announced in January 2020 demonstrates the government's drive for innovation. But clearer national leadership is needed. Though there is strong momentum in data ethics in policing at a national level, the picture is fragmented with multiple governance and regulatory actors, and no single body fully empowered or resourced to take ownership. The use of data analytics tools in policing carries significant risk. Without sufficient care, processes can lead to

outcomes that are biased against particular groups, or systematically unfair. In many scenarios where these tools are helpful, there is still an important balance to be struck between automated decision-making and the application of professional judgement and discretion. Given the sensitivities in this area it is not sufficient for care to be taken internally to consider these issues; it is also critical that police forces are transparent in how such tools are being used to maintain public trust.

In **local government,** we found an increased use of data to inform decision-making across a wide range of services. Whilst most tools are still in the early phase of deployment, there is an increasing demand for sophisticated predictive technologies to support more efficient and targeted services.

By bringing together multiple data sources, or representing existing data in new forms, data-driven technologies can guide decision-makers by providing a more contextualised picture of an individual's needs. Beyond decisions about individuals, these tools can help predict and map future service demands to ensure there is sufficient and sustainable resourcing for delivering important services. However, these technologies also come with significant risks. Evidence has shown that certain people are more likely to be overrepresented in data held by local authorities and this can then lead to biases in predictions and interventions. A related problem occurs when the number of people within a subgroup is small. Data used to make generalisations can result in disproportionately high error rates amongst minority groups.

Data-driven tools present genuine opportunities for local government. However, tools should not be considered a silver bullet for funding challenges and in some cases additional investment will be required to realise their potential. Moreover, we found that data infrastructure and data quality were significant barriers to developing and deploying data-driven tools effectively and responsibly. Investment in this area is needed before developing more advanced systems.

Sector-specific recommendations to regulators and government

Most of the recommendations in this report are cross-cutting, but we identified the following recommendations specific to individual sectors. More details are given in sector chapters below.

Recruitment:

- **Recommendation 1:** The **Equality and Human Rights Commission** should update its 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 **Information Commissioner's Office** 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.

Policing:

- **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.

Local government:

- **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.

Addressing the challenges

We found underlying challenges across the four sectors, and indeed other sectors where algorithmic decision-making is happening. In Part III of this report, we focus on understanding these challenges, where the ecosystem has got to on addressing them, and the key next steps for organisations, regulators and government. The main areas considered are:

- The **enablers** needed by organisations building and deploying algorithmic decision-making tools to help them do this in a fair way (see Chapter 7).
- The **regulatory levers,** both formal and informal, needed to incentivise organisations to do this, and create a level playing field for ethical innovation (see Chapter 8).
- How the **public sector,** as a major developer and user of data-driven technology, can show leadership in this area through **transparency** (see Chapter 9).

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 expected norms.

We want a system that is fair and accountable; one that preserves, protects or improves fairness in decisions being made with the use of algorithms. **We want to address the obstacles that organisations may face to innovate ethically, to ensure the same or increased levels of accountability for these decisions and how society can identify and respond to bias in algorithmic decision-making processes.** We have considered the existing landscape of standards and laws in this area, and whether they are sufficient for our increasingly data-driven society.

To realise this vision we need clear mechanisms for safe access to data to test for bias; organisations that are able to make judgements based on data about bias; a skilled industry of third parties who can provide support and assurance, and regulators equipped to oversee and support their sectors and remits through this change.

Enabling fair innovation

We found that many organisations are aware of the risks of algorithmic bias, but are unsure how to address bias in practice.

There is no universal formulation or rule that can tell you an algorithm is fair. Organisations need to identify what fairness objectives they want to achieve and how they plan to do this. Sector bodies, regulators, standards bodies and the government have a key role in setting out clear guidelines on what is appropriate in different contexts; **getting this right is essential not only for avoiding bad practice, but for giving the clarity that enables good innovation.** However, all organisations need to be clear about their own accountability for getting it right. Whether an algorithm or a structured human process is being used to make a decision doesn't change an organisation's accountability.

Improving diversity across a range of roles involved in the development and deployment of algorithmic decision-making tools is an important part of protecting against bias. Government and industry efforts to improve this must continue, and need to show results.

Data is needed to monitor outcomes and identify bias, but data on protected characteristics is not available often enough. One reason for this is an incorrect belief that data protection law prevents collection or usage of this data. Indeed, there are a number of lawful bases in data protection legislation for using protected or special characteristic data when monitoring or addressing discrimination. But there are some other genuine challenges in collecting this data, and more innovative thinking is needed in this area; for example, around the potential for trusted third party intermediaries.

The machine learning community has developed multiple techniques to measure and mitigate algorithmic bias. Organisations should be encouraged to deploy methods that address 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 contexts.** There is insufficient legal clarity concerning novel techniques in this area. Many can be used legitimately, but care is needed to ensure that the application of some techniques does not cross into unlawful positive discrimination.

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 (ONS)** 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 (ONS)** 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.

Recommendations to regulators

- **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.

Good, anticipatory governance is crucial here. Many of the high profile cases of algorithmic bias could have been anticipated with careful evaluation and mitigation of the potential risks. Organisations need to make sure that the right capabilities and structures are in place to ensure that this happens both before algorithms are introduced into decision-making processes, and through their life. Doing this well requires understanding of, and empathy for, the expectations of those who are affected by decisions, which can often only be achieved through the right engagement with groups. Given the complexity of this area, **we expect to see a growing role for expert professional services** supporting organisations. Although the ecosystem needs to develop further, there is already plenty that organisations can and should be doing to get this right. Data Protection Impact Assessments and Equality Impact Assessments can help with structuring thinking and documenting the steps taken.

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.

The regulatory environment

Clear industry norms, and good, proportionate regulation, are key both for addressing risks of algorithmic bias, and for promoting a level playing field for ethical innovation to thrive.

The increased use of algorithmic decision-making 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 limited case law or statutory guidance directly addressing discrimination in algorithmic decision-making, and the ecosystems of guidance and support are at different maturity levels in different sectors.

Though there is only a limited amount of case law, the recent judgement of the Court of Appeal in relation to the usage of live facial recognition technology by South Wales Police seems likely to be significant. One of the grounds for successful appeal was that South Wales Police failed to adequately consider whether their trial could have a discriminatory impact, and specifically that they did not take reasonable steps to establish whether their facial recognition software contained biases related to race or sex. In doing so, the court found that they did not meet their obligations under the Public Sector Equality Duty, even though there was no evidence that this specific algorithm was biased. **This suggests a general duty for public sector organisations to take reasonable steps to consider any potential impact on equality upfront and to detect algorithmic bias on an ongoing basis.**

The current regulatory landscape for algorithmic decision-making consists of the Equality and Human Rights Commission (EHRC), the Information Commissioner's Office (ICO) and sector regulators. **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 the overlap between discrimination law, data protection law and sector regulations is becoming increasingly important. We see this overlap playing out in a number of contexts, including discussions around the use of protected characteristics data to measure and mitigate algorithmic bias, the lawful use of bias mitigation techniques, identifying new forms of bias beyond existing protected characteristics.**The first step in resolving these challenges should be to clarify the interpretation of the law as it stands,** particularly the Equality Act 2010, both to give certainty to organisations deploying algorithms and to ensure that existing individual rights are not eroded, and wider equality duties are met.

However, as use of algorithmic decision-making grows further, **we do foresee a future need to look again at the legislation itself,** which should be kept under consideration as guidance is developed and case law evolves.

Existing regulators need to adapt their enforcement to algorithmic decision-making, and provide guidance on how regulated bodies can maintain and demonstrate compliance in an algorithmic age. Some regulators require new capabilities to enable them to respond effectively 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. Many regulators are working hard to do this, and the ICO has shown leadership in this area both by starting to build a skills base to address these new challenges, and in convening other regulators to consider issues arising from AI. Deeper collaboration across the regulatory ecosystem is likely to be needed in future.

Existing regulators need to adapt their enforcement to algorithmic decision-making, and provide guidance on how regulated bodies can maintain and demonstrate compliance in an algorithmic age.

Outside of the formal regulatory environment, there is increasing awareness within the private sector of the demand for a **broader ecosystem of industry standards and professional services to help organisations address algorithmic bias.** There are a number of reasons for this: it is a highly specialised skill that not all organisations will be able to support, it will be important to have consistency in how the problem is addressed, and because regulatory standards in some sectors may require independent audit of systems. Elements of such an ecosystem might be licenced auditors or qualification standards for individuals with the necessary skills. Audit of bias is likely to form part of a broader approach to audit that might also cover issues such as robustness and explainability. Government, regulators, industry bodies and private industry will all play important roles in growing this ecosystem so that organisations are better equipped to make fair decisions.

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 technical 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.

Recommendations to regulators

- **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.

- **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.

Public sector transparency

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.

A range of transparency measures already exist around current public sector decision-making processes; both proactive sharing of information about how decisions are made, and reactive rights for citizens to request information on how decisions were made about them. **The UK government has shown leadership in setting out guidance on AI usage in the public sector, including a focus on techniques for explainability and transparency.**

However, more is needed to make transparency about public sector use of algorithmic decision-making the norm. There is a window of opportunity to ensure that we get this right as adoption starts to increase, but it is sometimes hard for individual government departments or other public sector organisations to be first in being transparent; a strong central drive for this is needed.

The development and delivery of an algorithmic decision-making tool will often include one or more suppliers, whether acting as technology suppliers or business process outsourcing providers. 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. Procurement processes should be updated in line with wider transparency commitments to ensure standards are not lost along the supply chain.

Next steps and future challenges

This review has considered a complex and rapidly evolving field. There is plenty to do across industry, regulators and government to manage the risks and maximise the benefits of algorithmic decision-making.

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

Outside of specific activities, and noting the complexity and range of the work needed across multiple sectors, we see a key need for national leadership and coordination to ensure continued focus and pace in addressing these challenges across sectors. This is a rapidly moving area.

A level of coordination and monitoring will be needed to assess how organisations building and using algorithmic decision-making tools are responding to the challenges highlighted in this report, and to the proposed new guidance from regulators and government. 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 specific regulator or in CDEI.

In this review we have concluded that there is significant scope to address the risks posed by bias in algorithmic decision-making within the law as it stands, but if this does not succeed then there is a clear possibility that future legislation may be required. We encourage organisations to respond to this challenge; to innovate responsibly and think through the implications for individuals and society at large as they do so.

Recommendations to government

- **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.