Using AI to make or influence workplace decisions: legal considerations and mitigating risk

Moira Campbell, Director, at Fieldfisher LLP considers the risks of using AI in the Workplace, and how best to mitigate them

Employers are increasingly using AI-enabled tools in the workplace to automate and simplify existing processes. Whilst applying objective and data-based decision making to workforces can improve efficiency and consistency, using AI tools to make or inform decisions about applicants and employees is not without legal risk. Data protection, discrimination and breach of trust and confidence issues can arise, which may escalate into legal disputes. This article looks at the legal framework in the UK and the risks around using AI for workplace decisions, and the mitigation of those risks.

What is AI and how can it be used in the workplace?

AI systems aim to mimic human intelligence so that they can undertake tasks that otherwise would require human input. The OECD (an intergovernmental organisation focusing on economic progress and world trade) defines AI as: "a machine-based system that for explicit or implicit objectives, infers from the inputs it receives, how to generate outputs such as predictions, content, recommendations, or decisions, that can influence physical or virtual environments|".

AI can be used by employers throughout the employment lifecycle, for example:

- to devise job adverts, source candidates, screen CVs, and interview and select candidates during the recruitment process;
- to assess performance, by tracking employees to monitor productivity or health and safety in the workplace;
- to performance manage or allocate tasks to employees, including scheduling shifts and evaluating performance; and
- to discipline and dismiss staff, resulting from a decision or a score produced by the AI tool.

The impact of the law - existing relevant legislation

To date, the UK government has adopted a 'pro-innovation' stance to regulating AI. Instead of immediate legislation, it has pursued a principles-based approach meaning that there is currently no explicit AI-regulating legislation. The UK proposes a contextual, sector-based regulatory framework to leverage existing regulators and laws to govern AI efficiently. The use of AI in the workplace must therefore be assessed in the context of existing legislation.

Equality Act 2010

The Equality Act 2010 provides various protections against discrimination, such as protection against direct and indirect discrimination, harassment, and victimisation on the grounds of any of the nine protected characteristics (age; disability; gender reassignment; marriage and civil partnership; pregnancy and maternity; race; religion or belief; sex; sexual orientation). This protection from discrimination extends to candidates who are not yet employees but who are applying for roles, so it will also apply in the recruitment context.

The outputs produced by AI tools are not guaranteed to be objective and free from bias, and therefore are not guaranteed to be non-discriminatory. In most cases, machine learning systems will learn from the data input it receives, so any bias, skew, or discrimination in the input data will affect the output. In the employment context, to decrease the risk of discrimination claims and increase the reliability of the system, training and testing data needs to be comprehensive and representative of the real-world demographic of the employer's workplace. Otherwise, the input can inadvertently introduce and amplify existing hidden human biases.

This in turn can lead to discrimination, such as the widely reported Amazon CV screening algorithm, which preferred male candidates and penalised women, due to

an inherent bias in a decade's worth of internal recruitment data.

Health & Safety

When implementing a workplace AI system, it is important to consider the potential impact on employees. For example, if a particular monitoring tool would be likely to put 'constant pressure' on an employee (such as meaning they had to justify every short break), then consider how the processing can be minimised. Extremely close and precise monitoring could lead to employee sickness absence and health and safety issues such as complaints relating to workplace stress. The Amazon France Logistique CNIL decision is an example of excessive monitoring at work. In that case, all employee work breaks/interruptions were monitored, as well as their speed of work, and the details transmitted in real time to their line manager.

> ***"The UK proposes a contextual, sector-based regulatory framework to leverage existing regulators and laws to govern AI efficiently. The use of AI in the workplace must therefore be assessed in the context of existing legislation"***

Constructive unfair dismissal

In England and Wales, the term of 'mutual trust and confidence' is implied into all employment contracts. This means that the employer must not, without reasonable and proper cause, conduct itself in a manner calculated or likely to destroy or seriously damage the relationship of trust and confidence between the employer and employee. An employee could seek to claim that AI tools which put them under constant pressure and/or operate without transparency constitute a breach of trust and confidence, or that substituting an AI decision for an employer's own judgment legally undermines the basis of the employment contract. In these circumstances an employee may be able to resign and claim 'constructive unfair dismissal'.

GDPR compliance

Employers are acutely aware of the wide reach of the General Data Protection Regulation (GDPR). Employers will be the controller of any employee personal data processed as an input or output of AI tools in the recruitment or workplace context, meaning that the full suite of GDPR controller obligations will apply. Before using an AI tool in the workplace, employers must therefore identify a legal basis for processing any personal data and ensure that the processing is fair.

The GDPR requires that employers are clear, open, and honest with people about how and why they use personal data. This in turn requires employers to be transparent about how and why any AI assisted decisions about employees were made, or how their personal data was used to train and test an AI system. Employees also need to know how the AI is being used in order to enable them to request reasonable adjustments where necessary.

However, machine learning models generate their results by operating on high dimensional correlations that are often beyond the interpretative capabilities of human scale reasoning. In these situations, the rationale of algorithmically produced outcomes that directly impact employees remains opaque to both employees and employers. When the output could be discriminatory or unfair, the opaqueness of the model is problematic.

"Special category data", such as race, ethnicity and biometric data (where used for the purpose of identification) attract additional protections under the GDPR. Employers will need to have a clear understanding of if and how any special category data or biometric data is being used for machine learning within the business. For example, race and ethnicity data is probably needed to ensure that the model does not produce any discriminatory outputs, and also to ensure that training data is sufficiently representative. However, employers must be mindful of the need to establish a legal basis for processing that sensitive data.

Beyond the familiar requirements of the GDPR, under Art 22 of the GDPR an individual also has the right not to be subject to a "decision based solely on automated processing" which produces legal effects for that individual or similarly significantly affects them (e.g termination of employment). If the automated decision-making falls within the scope of Art 22, explicit consent of individuals is required, unless the processing is authorised by national law. It also invites additional transparency obligations, (for example, employers must be able to explain the logic involved in the automated processing), and additional data subject rights, such as the right for employees to contest the decision and obtain a human review.

Employees have the legal right to formally request information about, and access to, the personal data that an organisation holds about them by making a Data Subject Access Request. It is therefore important for employers who are users of an AI system to consider how well the system can respond to these requests. Consider, for example, ensuring that the datasets are searchable and can facilitate rights requests.

Upcoming legislation

There are a number of forthcoming pieces of legislation and related guidance that have been specifically drafted to regulate the use of AI in the business and employment environment.

(Continued on page 4)

(Continued from page 3)

EU AI Act

While many jurisdictions have taken a similar approach to the UK, in March 2024 the EU approved the EU Artificial Intelligence Act, which is anticipated to become the international standard and is another example of the so-called 'Brussels-effect'. The EU AI Act will take effect two years after it comes into force and will apply to both public and private organisations inside and outside the EU, as long as the AI system is placed on the market in the EU, or its use affects people located in the EU.

The Act categorises different technologies based on their level of risk, ranging from prohibited (such as biometric categorisation tools that categorise natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation) – which would see the technology banned – to high-risk systems and limited risk systems.

The EU AI Act classifies AI used in the workplace, including tools used for reviewing applications and making decisions in the employment life cycle, as 'high risk'. This will have significant consequences for employers deploying these tools, as there will be a range of strict compliance requirements including governance mechanisms, risk assessments, registration, human oversight, record keeping and increased transparency obligations.

Current and anticipated guidance

There is a wealth of guidance issued by regulators on AI. For example, the French CNIL has published its "AI how-to sheets" on the development of AI systems and creation of training datasets. We have also seen non-binding, voluntary codes and frameworks on how AI should be adopted. These include the OECD Principles (published in 2019), which was the first intergovernmental attempt to provide a framework for the responsible stewardship of AI.

DSIT

In March 2024, the Department for Science, Innovation and Technology ('DSIT') issued its 'Responsible AI in Recruitment Guidance'. As well as guidance on the set-up of systems, this highlights some of the key risks associated with deploying AI-enabled tools in recruitment processes. It sets out processes, assurance measures and mechanisms for employers to consider and put in place before, during and after AI procurement and deployment in the workplace.

EHRC

On 30 April 2024, the Equality and Human Rights Commission ('EHRC') published an update on its 'approach to regulating AI. The EHRC has prioritised AI within its strategic plan for 2022 – 2025. It will be focusing on reducing and preventing digital exclusion in 2024–25, particularly for older and disabled people.

The EHRC has voiced concern about the use of facial recognition technology ('FRT'). It supported the Claimant in the recently settled FRT Employment Tribunal discrimination case of *Manjang v Uber Eats UK Limited*, in which Mr Manjang alleged that ethnically biased facial recognition technology led to his suspension. Whilst this case did not ultimately test whether using FRT was discriminatory, it demonstrated that there is significant public interest in the potential for discrimination by AI in the workplace. We can expect continued interest in this area.

ICO

On 1 May 2024, the Information Commissioner's Office ('ICO') published "Regulating AI: The ICO's strategic approach". It explains the steps that the ICO is taking to drive forward the principles set out in the AI Regulation White Paper. This includes providing, for example:

1. Guidance on AI and data protection, automated decision-making and profiling, explaining decisions made with AI and biometric recognition technologies;

2. Advice and support for AI innovators, including the ICO Regulatory Sandbox and Innovation Hub services;

3. Enforcement action, which can include issuing information notices, assessment notices, enforcement notices and monetary penalty notices; and

4. Collaboration with other regulators, the government, standards bodies and international partners.

Consultations are planned to gather input on updates to ICO guidance on AI and data protection and automated decision-making and profiling in spring 2025. There will also be continued focus on biometric technologies.

Proposed legislation

There have been a number of proposals for AI legislation in the UK, which we may see getting pushed back following the general election.

TUC AI (Employment and Regulation) Bill and UK AI Bill

In April 2024, the TUC published the draft Artificial Intelligence (Employment and Regulation) Bill which sets out a potential UK legislative framework for regulating the use of AI in the workplace. It aligns with the EU AI Act, by focusing on 'high-risk' AI decisions relating to employment matters. 'High-risk' is widely defined, broadly capturing a decision that could impact workers' legal rights or significant aspects of employment. This could include hiring, firing, and/or assessing performance, for example.

This proposed legislation also competes with another UK Artificial Intelligence (Regulation) Bill, a private members' bill that, prior to the general election, received a third reading in the House of Lords. This private members' bill is based on principles-based regulation of AI.

Suffice to say that the precise form of law that will be enacted in this area, if any, remains uncertain.

Summary of risks

The key risks arising from breach of the applicable legislation and guidance can be summarised broadly as: individual and/or collective employment complaints, regulatory intervention, legal disputes, reputational damage, financial penalties and/or compensation and/or enforcement action for privacy violations. There is also always the risk that large language models may hallucinate, resulting in inaccuracy and/or bias.

Practical steps to mitigate risk

Governance is an ongoing responsibility throughout all stages of the AI lifecycle.

An effective **AI Governance Framework** will include:

- Clear policies, including specific AI policies, setting out how AI will be embedded in the organisation. Existing policies such as code of conduct, acceptable use of IT, data protection and privacy policies should be updated to refer to the use of AI tools.
- Accountability structures and processes should be implemented to ensure appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks.

Pre-deployment action:

- Consult with employees who will use the system to understand what training or skills they may require. Noting that under the EU AI Act workers and their representatives must be informed that they are subject to an AI system.
- Conduct due diligence of suppliers including asking them to provide 1) a 'model card' – a standardised reporting tool for capturing key facts about AI models; and 2) information about what data has been used to train the model.
- Train users on the intended purpose of the system, existing rules and regulations, and how to use AI tools and the data appropriately and securely.
- Implement Algorithmic Impact Assessments, Equality Impact Assessments and bias audits. For example, review the system for non-inclusive language and ensure that it has been tested on a wide demographic.
- Plan reasonable adjustments to remove disadvantages to disabled employees. If a reasonable adjustment cannot be made to remove the disadvantage to the disabled person, it may require the AI system's removal from the workplace.
- In many cases, AI will result in "high risk" to candidates and employees and will trigger a legal obligation to conduct a Data Protection Impact Assessment (DPIA). This helps to show accountability, by documenting the risks to individuals and any mitigation measures taken, a clear purpose and lawful basis for the use case. The DPIA is important because it will be one of the first pieces of documentation a privacy regulator will ask to see if there is a complaint against the organisation or an investigation.
- Conduct a thorough pilot, which will not only ensure that employees understand how the system works, but also might reveal any bias or inaccuracy that had not yet been detected.

Live operation:

- Monitor the systems by repeating bias audits and performance testing every 6 months from when the AI system goes live and/or when major new updates are released, to ensure that they are performing as intended and not producing discriminatory, erroneous or unjustified results.
- Implement a user feedback system.
- Maintain human oversight of the system and its outputs, so that decisions are not solely automated/ there is the option of a human explaining the decision, and to guard against hallucinations.
- Ensure there is the option of contestability to enable individuals to challenge the decision and obtain human intervention.

AI presents a myriad of compliance challenges within an evolving area of law. The current landscape patches together existing legislation, regulation, guidance and frameworks, from which we can extrapolate some core principles to give us a framework to build on. Managing AI is likely to be an area where there are increased calls for legislation and it is likely to be a future focus for government.

Moira Campbell
Fieldfisher LLP
moira.campbell@fieldfisher.com