Day 4
by Paul Schwartfeger

When AI Tips the Scales: Bias and Unfairness

AI systems are moving rapidly from conversation into decision-making, from ranking CVs to allocating public services. The efficiency is tempting, though bias in training data or configuration can turn automation into unlawful discrimination.

In 2018, Amazon abandoned its AI internal recruitment tool as it reportedly downgraded CVs containing terms linked to women's colleges or women's activities. The tool had "learned" to do so from historical hiring data skewed toward men.

Applicable legal principles
Under the Equality Act 2010, discrimination may be direct under s.13, where less favourable treatment arises because of a protected characteristic, or indirect under s.19, if a provision, criterion or practice is applied to all though it disproportionately disadvantages a protected group without objective justification.

From a GDPR perspective, automated decisions producing legal or similarly significant effects engage Articles 5(1)(a) and 22. The principle of "fairness" means decisions must not be based on biased or irrelevant factors. Article 22 restricts sole reliance on automated decision-making without meaningful human review.

However, AI can complicate claims under equality and data protection laws. Attribution is one problem. Was the bias present in the base model, any fine-tuning, or caused by a user's prompt?

The "black box" nature of models also makes it difficult to show how a protected characteristic influenced an output, creating evidential challenges.

In actions for direct discrimination, identifying a valid comparator can be difficult. Outputs from generative models may vary between operations even for identical inputs, making it hard to produce a like-for-like comparison. Many deployed systems also personalise or adapt responses based on user-specific data, so the "treatment" given to one individual is not easily replicated for another. Without disclosure of system logic and datasets, establishing that a claimant was treated less favourably than a real or hypothetical comparator may be evidentially complex.

Potential defendants
Depending on the facts, liability could fall on an employer in a recruitment context, a system developer for failing to address bias in underlying logic, or contractors whose configuration choices embed discrimination. End users who knowingly use a biased system could also be at risk.

The extraterritorial reach of the GDPR means that liability can still arise where individuals in the UK or EU are affected, even if the provider is overseas.

While not binding in the UK, the decision by the CJEU in *SCHUFA (C-634/21)* shows how liability for decisions based on automated processing can extend "upstream" to those involved in preparatory acts, even if a different party delivers the final decision.

Evidence and interim remedies
Key evidence may include logs showing how the AI tool responded, configuration settings and statistical analysis of disparate impact, similar to the comparative approach in *Enderby v Frenchay Health Authority* [1993] ICR 112. Disclosure will also often be essential to uncovering bias in the underlying training data or algorithms.

Interim relief, such as orders to preserve evidence, suspend the use of a disputed model or prevent reliance on particular outputs may aso be needed to avoid further harm before trial.

Potential defences
Under the Equality Act, defendants may deny any causal link between outputs and a protected characteristic, or in an indirect discrimination claim argue that the system's approach is a proportionate means of achieving a legitimate aim.

Under the GDPR, defendants may claim decisions were not solely automated, lawful bases existed, or bias was minimal and addressed by safeguards. Claimants might counter with proof of systemic disparities and a lack of meaningful human review.

Remedies
Remedies under the Equality Act include compensation for financial loss and injury, declarations and recommendations for change. Under the GDPR, individuals may claim compensation for material or non-material damage (Article 82) and seek orders to halt unlawful processing.

Conclusions
Bias in AI-driven decisions is no longer just a compliance risk, it is a litigation risk under equality and data protection laws. The challenge for lawyers is to map the full chain from input to output and to secure evidence before it disappears.

Paul Schwartfeger

Paul is a barrister with 36 Stone, specialising in commercial litigation and international arbitration with a particular focus on legal disputes involving data and technology.

psc@36stone.co.uk

+44 (0) 20 7440 6900