



AI and the Rebirth of Mass Data Claims

JANUARY 2025

Representative group action was dead; to begin with. Recent judgments have demonstrated the difficulties for claimants in sustaining mass opt-out claims against tech giants for data misuse. However, AI may offer a solution by enabling the individualised assessment of damages claims at scale.

There is an inherent tension between the importance of doing individual justice to a particular claimant, and the efficiency gains to be obtained from managing many similar claims together.

Under CPR 19.8 a claim may be pursued on an opt-out basis by a representative who has the same interest as the other claimants. Any judgment is then binding on all persons represented. The difficulty lies in showing that the representative has the 'same interest' as the rest of the group.

The internet, social media, and now AI have given rise to many situations in which group action proceedings for data misuse might appear an attractive prospect for litigants.

A claim might arise as follows: a technology company provides a service. People engage with that service, in so doing providing user data to the company. The data is then utilised to augment, improve and create new technologies from which the company can profit. The problem comes where the company processes the data in a way which goes beyond the scope of lawful use. Data hungry AI models exacerbate the need for that data to be sourced for their training.

For data breach claimants, individual damages are likely to be relatively modest; it is only if those claims are aggregated that the costs of pursuing litigation are likely to be worthwhile.

In recent years, the appellate courts have twice engaged with representative claims for data misuse, with both cases resulting in victories for tech companies.

AUTHOR



Jacob Turner
Call Date: 2016

[Read more](#)

In [*Lloyd v Google*](#) the Supreme Court considered allegations that Google had tracked the activity of millions of iPhone users without their consent. The claimants said that Google had breached its duties under the Data Protection Act 1998.

Mr Lloyd needed to show that there was a 'lowest common denominator' to the entire class of claimants. The Supreme Court held that this lowest common denominator would have been an iPhone user who had clicked on a relevant website on a single occasion but had received no targeted advertisements as a result. Those facts did not overcome the seriousness threshold which applied to data protection claims as it was impossible to characterise such damage as more than trivial.

Mr Lloyd argued therefore that a 'uniform sum' of damages should be awarded in relation to each person whose data protection rights have been infringed without the need to investigate any individual case. The Supreme Court rejected this proposal: the mere fact of data misuse did not give rise to damages. Each member of the class needed to have a realistic prospect of success. Mr Lloyd's representative claim was therefore refused permission to proceed.

Recently, in [*Prismall v Google*](#), the Court of Appeal dealt another blow to representative claims. The claimants sought damages under the tort of misuse of private information for transfers by a hospital of patient medical data to Google and DeepMind. Again the appeal depended on whether there was a lowest common denominator of harm between all of the 1.6 million potential claimants. The Court of Appeal held that there was no irreducible minimum harm because it could not be said that the claimants had any shared expectations of privacy. Indeed

some within the class of claimants had publicly shared their medical data on social media. Mr Prismall's claim was struck out.

In early 2025 the High Court handed down a further judgment which reiterated the difficulties of satisfying the CPR 19.8 test. In [*Getty Images \(US\) Inc & Ors v Stability AI Ltd*](#) the Court rejected an application for the certification of a representative action. In that case an intellectual property rights holder, Mr Barwick, sought to bring a claim on behalf of a wider group of parties whose property was said to have been infringed in the training process of an AI program undertaken by the defendant, Stability. The Court noted that the exercise of inviting the claimants to identify the core propositions of the case that they were seeking to advance exposed the extent to which a proposition at the very high level of generality proposed by the claimants was incapable of satisfying the requirement that all members of the proposed class have the same interest.

The outcomes in *Lloyd, Prismall* and now also *Getty* emphasise just how difficult it is for a representative data misuse claim to succeed. The damages are likely to be related to distress or inconvenience, and their assessment seems to depend on circumstances individual to each litigant.

In addition to CPR 19.8 there are several ways in which civil claims can be advanced on behalf of a group of claimants but none of them offer a particularly attractive solution for data misuse claims.

One route is via a 'group litigation order'. GLOs typically involve a large number of small claims, which enable economies of scale for those managing the litigation. These claims are opt-in because each individual claimant must authorise their inclusion in the group. As a result, they will

typically require a significant initial investment by any funders in building the class of claimants.

In 2009 the Government rejected a recommendation from the Civil Justice Council to introduce a generic class action regime, preferring instead a 'sector based approach'. To date, competition law is the only sector for which a specialised 'class action' regime has been introduced. Under the Consumer Rights Act 2015, there are two key procedural advantages for competition law claimants. First, opt-out proceedings enable members of the class to have their claims advanced without their knowledge. Second, they can be awarded aggregate damages, without the need for proof of loss by individual claimants.

The attractive aspects of competition actions may encourage claimant groups to shoehorn their causes of action into competition law ones, but plainly this will not be possible for every claim. The procedural relaxations for competition claims highlight the difficulties under the unreformed CPR 19.8. If the current trend continues there may be pressure on the Government to introduce targeted reforms to data litigation, but none appear imminent, and such reforms will likely face trenchant opposition from the major technology companies.

There may be another way through.

An AI system could be trained on the data of claimants to provide an individualised rating and valuation for the damage caused by the relevant data misuse. The lowest common denominator between such claimants need not therefore be an arbitrary figure of the type rejected by the Supreme Court in *Lloyd*.

How would the AI system work? The technology involved would likely comprise a purpose-built 'discriminative classifier', meaning that the system sorts and scores existing data, rather than a 'generative model', which creates new output – such as ChatGPT. Consequently, it is unlikely that the AI systems envisaged here would be subject to the type of hallucination errors for which generative AI is known. Any AI system is of course subject to errors, but those of discriminative classifiers are typically capable of being measured, quantified and reduced to an acceptable level.

There would need to be relatively uniform data for each claimant, relevant to the harm said to have been suffered. Fortunately, for mass claims arising from data misuse, this step ought not to be particularly difficult: the misused data which is subject to the claim could provide the training data for the claimants' own AI system. Such data is presumably of at least some predictive value – otherwise the tech companies would not have taken and used it in the first place. For the initial group of claimants that data could be obtained at no cost via a Subject Access or Data Portability Request under the GDPR.

Once the claimant data has been gathered, the AI system would need to be trained on different types of claimant in order for the gravity and effect of misuse to be calibrated. A human-led assessment and scoring metric could be applied to a sample of the claimant data, which might then be used to create a model capable of predicting values for new, unseen data.

Finally it would be necessary to plead a representative claim with sufficient specificity so as to include only the class of claimants whose claims overcome a scoring threshold established by the

deployers of the AI system, whether in terms of the sustainability of the cause of action, viability under CPR 19.8, or for any higher economic threshold determined by the group's funders. Courts have already accepted the possibility of statistical methods being used to make good crucial aspects of claims. In [*Josiya v BAT*](#), the High Court refused an application for strike out of a business and human rights claim, in circumstances where a model of mathematical probability was said to link over 7,000 claimants, via their harvesting of tobacco leaves, to purchases made by particular defendants.

Clearly there would need to be an initial investment in creating the AI system and building a sufficient initial class of known claimants on whose data the model could be trained. Once that investment has been made, the cost of scaling the system across new members of the potential claimant class would likely be minimal.

This approach will not be suitable for some fact patterns, particularly those where claimants stand to be excluded due to matters that cannot readily be modelled. In appropriate cases the use of AI could provide a way of cutting through the Gordian knot imposed by the current case law. It might work particularly well in the bifurcated procedure contemplated by Lord Leggatt in *Lloyd*, where liability and quantum in a representative claim are split: common issues of fact and law could be determined through a representative claim, leaving issues that require individual determination, whether relating to liability or damages to be dealt with at a subsequent stage. Whereas the cost of individual determinations might in the past have prevented bifurcated claims on economic grounds, the use of AI could

radically alter the equation.

We generate a wealth of data each time we use our phone or web browser, enabling tech companies to build extraordinarily detailed profiles, developing insights about aspects of our behaviours that we may not have known ourselves. The data collected is the product of billions of dollars of investment. Why not leverage that same information to facilitate legal claims where the data is misused?

There is precedent for AI systems being employed by claimants to generate evidence of wrongdoing and harm: in [*Allianz Global Investors & Others v Deutsche Bank & Others*](#), a multibillion dollar follow-on damages claim in the Competition Appeals Tribunal arising from an alleged forex cartel, the claimants alleged that they had suffered losses as a result of exchange rates being fixed by leading banks. This would have involved demonstrating wrongdoing, causation and harm across millions of individual trades. Possibly for the first time in any civil litigation, the claimants sought to plead and prove their case using an AI system. They claimed to have trained a machine learning algorithm to recognise forex fixing based on around 500 alleged 'known instances' of wrongdoing. The author of this article acted for a group of the defendant banks. The claim settled in early 2023 so we do not know what the outcome would have been but as a proof of concept it remains important, and could well be of wider significance.

Ironically, the same AI technology which is now frequently the cause of mass data misuse may also provide a solution for litigants.