Technology and Data:
Analysing the relationship between the power couple of AI-related research

03

Darryn Hale
Partner
dahale@dacbeachcroft.com

Christopher Air
Partner
cair@dacbeachcroft.com

Modern society loves power couples – from Posh and Becks through to Will and Kate, we are obsessed with glamorous celebrity couples who are collectively more successful than the sum of their parts and who are indispensable to each other's success, fame and happiness. Technology and data are the power couple of AI-related research – never out of the spotlight, the subject of endless discussion and often the source of controversy (but no paparazzi thankfully).

From training machine learning models to refining algorithms for natural language processing and computer vision, data plays a hugely influential role in determining the performance and reliability of AI applications. In research contexts, data is not merely a passive input; it is actively curated, pre-processed, and analysed to uncover patterns that drive AI innovation. However, the relationship between the commercial drivers for undertaking AI related research, and the legal limitations around using data for such purposes, have something of an uncomfortable history under UK law (not quite Meghan and Harry but you get the idea!). Given the clear desire for the nation to be at the forefront of AI discovery, the question arises as to whether our current legal framework will be made more flexible and accommodating for conducting AI research or whether counter pressures such as needing to maintain adequacy in the eyes of the European Commission for unhindered data flows with the EU result in us leaving UK data protection law largely unchanged in this respect? We therefore reflect on recent changes to UK data protection law made by the Data (Use and Access) Act 2025 (**DUAA**), as those changes specifically relate to research.

We have considered the legal environment in which our power couple operate, providing an overview of some of the legal and ethical challenges relating to use of data, particularly personal data, for AI related research, and explain the current legal position under UK data protection law in this context, which includes a look at the accountability framework of the Information Commissioner's Office (**ICO**).

We have also reflected on sources of flare ups and spats which our couple face daily; tensions that can arise between contracting parties as to their roles as either joint controllers, controllers and/or processors and wider concerns regarding supply chain risks. We then conclude by setting out some recommendations for managing such contractual tensions in practice – a guidebook for our power couple to live happily ever after!

AI-related research – the Legal and Ethical Challenges

Let's start by looking briefly at a few examples of some of the legal and ethical challenges posed by use of data for AI research.

- **Privacy:** AI systems often rely on large datasets containing personal data and even special category data. The party designated as the data controller will need to ensure that such processing fully complies with the UK General Data Protection Regulation (**UK GDPR**), Data Protection Act 2018 (**DPA 2018**) and relevant ICO guidance. Of course, fully anonymised data falls outside the remit of such legislation and we explore the concept of anonymisation in this collection.
- **Bias and Fairness:** If datasets are non-representative or skewed, AI models can perpetuate or amplify social biases, leading to discriminatory outcomes. This is particularly problematic in areas like hiring, lending, or law enforcement, where biased algorithms can harm marginalised groups.
- **Transparency and Accountability:** Many AI systems operate as 'black boxes,' making it difficult to explain decisions. Lack of transparency undermines trust and accountability, especially when AI influences critical decisions in healthcare, finance, or criminal justice for instance.
- **Data Ownership and Commodification:** Increasingly, personal data is treated as a commodity, raising ethical questions about ownership and exploitation. Individuals often have little control over how their data is monetised or shared.

A UK data protection law perspective

The UK GDPR and DPA 2018 set out specific provisions relating to the use of personal data for research purposes, breaking research down into the following types: (i) archiving purposes in the public interest; (ii) scientific or historical research purposes; and (iii) statistical purposes. Helpfully, ICO guidance explicitly lists the development of AI as falling within the scope of what constitutes 'research'. Furthermore, the DUAA has also introduced a new statutory definition of "scientific research" that covers any research reasonably described as scientific, whether publicly or privately funded, and whether commercial or non-commercial. The research provisions within the UK GDPR and DPA 2018 are scattered across several far flung parts of the legislation and require a reader to cross refer to, and understand the interplay between, data protection principals in Article 5 UK GDPR, data subject rights under Chapter 3 UK GDPR and safeguards under Article 89 UK GDPR, as well as exemptions under Schedule 2, DPA 2018. Understandably therefore, the research provisions are notoriously difficult to understand and apply in practice (at least with a sufficient degree of certainty).

Below we summarise the key provisions as they currently stand and look forward to how these are set to change by virtue of the DUAA.

- **Lawful Basis for Processing:** Research involving personal data must have a lawful basis under Article 6 UK GDPR, the most common being "public task", "legitimate interests" or "consent". Moreover, as data used for research is often special category data (e.g. information relating to health) a condition for processing is also required under Article 9 UK GDPR or Schedule 1, DPA 2018. One such condition is research related purposes under Article 9 UK GDPR, which is expanded upon under Schedule 1, DPA 2018. It restricts use of the data for measures or decisions about particular people, except for approved medical research; and requires that the processing be in the public interest. Furthermore, the DUAA introduced amendments allowing "broad consent" for processing personal data for scientific research purposes. This means consent remains valid even if specific research purposes cannot be fully identified at the time of collection, provided ethical standards are followed and participants can consent to parts of the research where possible.
- **Purpose Limitation & Compatibility:** Usually, personal data must only be processed for the original purpose for which it was collected. However, the UK GDPR allows further processing for scientific or historical research or statistical purposes if appropriate safeguards are in place in accordance with Article 89. These safeguards focus in particular on measures to ensure respect for the principle of data minimisation, including where possible, anonymising or pseudonymising data. The DUAA introduced some welcome flexibility around the purpose limitation principle, enabling reuse of personal data for research without requiring a new lawful basis, as long as safeguards are maintained. This is particularly helpful for secondary research using existing datasets.
- **Exemptions:** Schedules 2–4, DPA 2018 set out exemptions which apply to the use of personal data for research purposes. Such research related uses are exempt from certain data subject rights (e.g. access, erasure, objection) if applying those rights would seriously impair or prevent the research, and the required Article 89 UK GDPR safeguards are in place.
- **Storage Limitation:** Normally, personal data should not be kept longer than necessary. However, for research, archiving, or statistical purposes, data can be retained indefinitely if the Article 89 UK GDPR safeguards are in place.
- **Transparency:** Despite the challenges around explainability of AI systems and their outputs, transparency remains an important principle under Article 5 UK GDPR, so organisations will need to provide information to data subjects, informing them that their data will be used in AI related research and the resulting decisions. Indeed, the ICO has published specific guidance on this topic. However, the DUAA disapplies certain transparency requirements. Ordinarily, if using data for a purpose other than the original purpose, the controller needs to inform affected data subjects. However, under the DUAA this does not apply to scientific research if doing so is impossible or would involve disproportionate effort.

The DUAA therefore goes some way towards relaxing the rules around research. In terms of how an organisation complies with, and demonstrates its compliance with, these legal requirements, specifically in the context of AI research, it is useful to turn to the ICO's accountability framework, which we explore next.

A regulatory perspective

The ICO has produced an extensive accountability framework which is supplemented by specific guidance on the accountability and governance implications of AI. In essence, accountability is a matter of actively demonstrating and evidencing how the above legal compliance requirements have been met. Neither the framework nor the guidance necessarily focuses in on the implications of AI-related research, but nonetheless the principles set out are of sufficiently broad scope and effect to be relevant. The documents are extensive; headline points include:

1. A Data Protection Impact Assessment DPIA is required for any processing of personal data which involves the use of new technologies including AI, machine learning and deep learning.
2. There is a tension between data minimisation and statistical accuracy which, on the face of it, can be difficult to reconcile. In particular, the fairness principle under the UK GDPR requires any AI to be statistically accurate but at the same time the AI needs lots of data to properly train which potentially conflicts with data minimisation requirements. Ultimately this comes down to technical analysis of the minimum dataset required to achieve sufficient accuracy. This may need to be kept under review, particularly once AI moves beyond research into live deployment (i.e. checking whether the outputs remain within expected accuracy levels).
3. Careful consideration needs to be given to the nature of any training data used to research or develop AI, both to ensure that it is not imbalanced but also to correct for any past discrimination inherent in the dataset. This is to ensure compliance both with the fairness requirement under the UK GDPR but also broader anti-discrimination laws and may include steps such as adding or removing data relating to under/over-represented groups from the dataset.

In addition, the ICO has also recently focussed on AI supply chain contract management, including that related to generative AI technology (which is often more nuanced). It recently conducted a five-part consultation to clarify how UK data protection law applies to generative AI. The series addressed five key areas, namely: (i) lawful basis for web scraping; (ii) purpose limitation; (iii) accuracy of training data and outputs; (iv) engineering individual rights into AI models; and (v) allocating controllership across the AI supply chain. One point of particular interest is that while the consultation itself does not indicate or settle a formal legal position, it does challenge the traditional position of AI developers only ever being processors. This is a point we consider further below in the context of the inevitable constraints which arise from a controller/processor relationship being adopted.

Contractual tensions

Research is one of the areas in which we see most variation in approaches to contracting and in particular the assessment and attribution of roles under the UK GDPR i.e. whether a party is acting as a controller, joint controller or processor. This, to some extent, reflects the different ways in which AI-related research may be conducted: for instance, it could be a standalone research project in which one party sponsors another to gather data on their behalf in relation to a specific hypothesis or issue. It could also be a collaborative arrangement in which several parties agree to pool their data in order to enrich the datasets they individually hold and they conduct research on that pooled data. Finally, research could also be an adjunct to the delivery of services – for instance, a tech supplier delivering a particular product or service to a customer and wanting to use the data they collect when delivering those services in order to research potential new products or services.

As most will know, the question of designation under the UK GDPR is a factual one i.e. looking at who does what with the data and then assessing whether that is indicative of a controller, joint controller or processor role. In our experience, however, that is often not interrogated as robustly as it ought to be and instead habit creeps in – in the health sector, for example, a supplier of a tech-based solution or platform is usually deemed a processor by default because this is felt to be a way of safeguarding highly sensitive patient data. This then, inevitably, leads to a very tightly restricted set of instructions to the supplier in the contract which, realistically, stifle any possibility for them to use the data for other innovative purposes (notably research).

It is not only controller-processor contacts which are prone to contractual constraints, whether express or implied, as controller-controller and joint controller arrangements are often so sparse with detail that it is completely unclear as to what is deemed legitimate use of data under the contract.

We consider some of the common dynamics and tensions commonly encountered in the context of the above scenarios below.

1. Controller to processor

This scenario most commonly arises where one party is providing services to another, which may themselves be based on AI. Equally they may not but the data processed in order to deliver the services is sufficiently interesting to mean that it could be used to research and develop AI in the future. In turn, one of two things will usually happen: first, the processor is expressly prohibited from using personal data processed on the controller's behalf for any purposes other than delivery or second, the possibility of research is not contemplated when the contract is agreed and so is not addressed one way or another.

Either scenario would, effectively, constrain the possibility of using the relevant data to research and develop AI. It can also be further compounded, in our experience, by tech suppliers contracting with customers on the terms they supply and ending up with hugely variable approaches to data protection compliance (and which, self-evidently, is very difficult to unpick).

2. Separate controller to controller

In a separate controller to controller relationship, the picture is arguably somewhat more straightforward than with controller to processor, but not always. Essentially, each party bears its own responsibility for UK GDPR compliance, and either agrees to share data for different/shared purposes or simply does their own thing. The degree of co-operation or assistance between the parties to aid the other's compliance is variable. Either way, responsibility for use of data for aspects such as AI research is usually borne by each party.

In principle, therefore, this is a more favourable scenario for the purpose of enabling AI-related research. However, the common mistake is often that the data sharing provisions between the controllers are so vague or, at the other end of the spectrum, so tight that it is unclear whether use of shared personal data for research into AI was contemplated.

3. Joint controller scenario

This applies where there is collaboration between two or more parties, often with pooling of data. Article 26 UK GDPR sets out high level requirements for this scenario, but these aren't prescriptive. In practice, this means agreeing which party is responsible for providing the fair processing notice and handling data subject rights requests, and which acts as point of contact for requests from data subjects.

Concluding thoughts

Considering the above, our take from a contractual perspective is that in order to enable AI-related research (i.e. our recipe for a happy, harmonious and lasting power couple relationship) it is imperative to:

(i) consider whether the purpose can be achieved through the use of anonymised data;

(ii) properly analyse roles under data protection law and in particular whether a tech supplier is genuinely always a data processor;

(iii) develop consistent contractual template terms which either reflect a controller designation or, where a processor, contain authorisation from the controller to enable research processing; and

(iv) have honest up front discussions before agreeing contracts about the potential AI-related research use cases.