

ICO consultation on Purpose limitation in the Generative AI lifecycle

The British Copyright Council (BCC) represents those who create, hold interests, or manage rights in literary, dramatic, musical, and artistic works. The following response has been developed with our membership which include professional associations, industry bodies and trade unions which collectively represent the voices of over 500,000 creators, spanning the creative industries.

These rights holders include many individual freelancers, sole traders, and SMEs, as well as larger corporations within the creative and cultural industries. Our members also include collecting societies which represent rights holders, and which provide licensed access to works of creativity. Our members, individual creators or corporations have been using AI for over a decade as a tool assisting their creative activities or business administration.

We have commented in detail to the first chapter of the ICO series of consultations on generative AI and data protection. We reiterate the main message asking for mutual respect from AI developers when processing personal data (and generally) of BCC members; only a level playing field for all parts of the value chain will enable the establishment of a successful market in which AI developers innovate and prosper in tandem with the creative sector and society overall. The centrepiece of a fair market is compliance with the legal framework, for our members this means mainly copyright (including anti circumvention of technological protection measures), trademark, data protection, privacy, non discrimination and contractual obligations. In the same way as in the first chapter of this series of consultations on generative AI , the legality of the processing and ability of individuals to exercise their data protection and privacy rights is the fundamental key consideration. For the purposes of this response, we refer to "AI Developers" to include AI developers training models, adapting models, deploying models and each separate organisation involved in the AI lifecycle. Personal Data of BCC creators includes data such as names, likeness, voice , as well as potentially sensitive category data.

- I. Chapter 1 - Generative AI [first call](#) for evidence: the lawful basis for web scraping to train generative AI models

What are the possible lawful bases for collecting training data?

As part of complying with the lawfulness principle of data protection, developers need to ensure their processing:

*(a) is not in breach of any laws; and
(b) has a valid lawful basis under UK GDPR.*

As we maintain in our submission to the first call for evidence: without express permission of the creator or rights holder AI developers processing training data by web scraping, usually in breach of the terms of service/ licence terms are breaching amongst others intellectual property laws (a). Furthermore, there is no valid lawful basis under UK GDPR justifying processing without permission (b) as the only potentially available basis identified by ICO (with which we agree) is the legitimate interest test, and it is impossible to see how the legitimate interests test could be satisfied given the extent to which this processing affects the interests, fundamental rights and freedoms of the individual creator. Our members are confronting this issue currently when dealing with deep fakes and other synthetic, unverified and unauthorised content.

More details in the BCC [submission](#).

We repeat our support for the ICO's recommendation that AI developers training models and indeed at all stages in the AI lifecycle, to be required to undertake a written DPIA and otherwise complete written records of processing and the basis for doing so.

We also note that application of the UK data protection framework under UK GDPR etc is required in order to access European and other markets under international requirements and bilateral agreements.

Given that in the absence of express consent for the processing to train models ,AI developers do not comply with the lawfulness principle of data protection in the first place in our view, the purpose limitation constitutes a secondary consideration. It is worth repeating that data processing by scraping the web to train generative AI models without express consent and compliance with the other key data protection principles infringes the lawfulness principle of data protection and fundamentally causes damage to individuals compounded by an absence of redress or ability to enforce their rights as individuals given they have not been informed of the web scraping. However it is nonetheless worth stating the fundamental point that if no consent or information has been given to the individual it is highly unlikely that any purpose limitation can have been considered or implemented by the original AI developer nor information provided to the individuals about the extent, type and nature of processing.

A specified, explicit and legitimate purpose

“Purpose limitation requires organisations to have a clear purpose for processing any personal data before they start processing it. If they are not clear about why personal data is processed, it follows they will not be able to be clear with individuals.

This purpose must be legitimate, meaning that:

- 1. There must be a lawful basis for processing it; and*
- 2. The purpose is not in breach of other laws, such as intellectual property or contract laws.*

We agree with the assessment by the ICO that AI developers need a specified, explicit and legitimate purpose in order to undertake data processing by web scraping. This purpose needs to be reviewed and reassessed for each distinct processing use of data within the AI lifecycle. In particular, if datasets are sold on, a new specified, explicit and legitimate purpose is required. In the area of generative AI, dataset providers i.e. the organisations scraping the web for training data, often sell them on to several AI developers. We imagine a scenario where a non-commercial dataset provider, arguing that at least as far as copyright is concerned they operate on a lawful basis, sells datasets to commercial organisations. Such activity evidently requires the re-establishment of a purpose limitation, re-evaluate the lawful basis for processing and, if intending to rely on the legitimate interests argument, that the purpose for processing itself is lawful. Notably, activities by dataset provider which they allege are non-commercial, do not remove responsibilities under contractual or anti-circumvention rules; they remain unlawful.

AI developers operating in the UK market need to comply with the lawfulness principle under UK data protection rules; if they want to prove compliance, they need to maintain comprehensive records of the data they are processing. In particular if they are relying on legitimate interest principles as established in the ICO [guidelines](#). Record keeping thus manifestly constitutes a key element of the required activities of dataset providers or AI developers intending to comply with UK data protection laws. As recognised by the ICO, this exercise must enable all parties, including the individuals whose data is used during the training and development of AI systems, to have a clear understanding of why and how the personal data is used in compliance with law.

In addition, as the ICO notes (and we fully support), the aim of data protection law is to empower individuals and give them greater control over their rights and fundamental

freedoms, including the right to be informed, rectify inaccurate or incomplete data and erasure. Individuals' rights, particular those of our members, are eroded the further their personal data is processed, adapted and distributed to models for deployment. In the absence of consent, information about what and how data is being used, individuals have no right of redress or even know who to contact in order to exercise their rights. We therefore support ICO in its' conclusion that it is imperative both that generative AI developers respect individual rights before they start processing and that they obtain consent from the relevant individual in relation to each stage and ensure data minimisation.

We broadly agree with the statements made by the ICO in the consultation paper. We are looking forward to working with the ICO as pertinent regulator protecting the personal as well as sensitive data of UK creators and artists.