

ICO consultation: Generative AI fourth call for evidence. Engineering individual rights into generative AI models

June 2024

Overview

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

Many BCC members are creators who increasingly work with AI technologies as assistive tools linked to the works they create. On the other hand, many creators are extremely concerned with good reason, that AI-outputs are, and will be used without recognition or permission from the human authors and the authors who make the arrangements for the creation of the works through the use of AI applications.

As such, transparency over how creative works and performances and the personal data associated with these works can be ingested and adapted throughout this process, particularly for works including personal data which are protected by copyright, will be increasingly important.

This can be accomplished by respecting existing UK copyright and data protection law frameworks. IP licensing safeguards will remain vital to protect against the unfair use and devaluation of copyright protected work.

Situation Analysis

It is currently the case that creative works and performances and the personal data associated with them are frequently being ingested for training generative AI applications without securing permission for copyright licensing or obtaining consent for the use of the personal data associated with the process.

The extraction and reproduction through text and data mining of creative works such as text, video, image and music from publicly available websites, even in cases where such practices are explicitly prohibited in their terms and conditions, has compounded the issue even further. Such creative works include personal data intrinsic to the individual creator/ performer and, depending on nature of the work, sensitive category data such as biometric data.

This has been recently exemplified by the unlicensed use of the “voice” of Scarlet Johansson by ChatGPT, who rejected a request for the use of her voice by ChatGPT.

The fact that creative works may be publicly accessible online for a specific authorised use does not mean they are "publicly available" for scraping when the controller of the data being scraped is potentially responsible for further use if the creator as no knowledge or control over what is being done.

Moreover, there are concerns that those who may argue that scraping or mining activities are justified under the exceptions or limitations for their own processing often fail to inform others (who remain responsible for the personal data) of what is being done and for what reasons. This leaves creative companies as “victims” liable to others under data processing rules when they themselves have no control or knowledge of how or why the personal data for which they are responsible has been taken and used.

There is an urgent need for such creative companies to be able to secure information from data scrapers to show that personal data processing rules are being observed and that unwitting liabilities are not attached to breaches by creative companies over which they have no control.

The solution: A level playing field

Only a level playing field for all parties in the value chain will ensure 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 and related rights (including anti circumvention of technological protection measures), trademark, data protection, privacy, non-discrimination and contractual obligations.

In the same way as in the first three chapters of this series of consultations on generative AI, the legality of the acquisition of data and subsequent processing and the ability of individuals to exercise their data protection and privacy rights is a fundamental consideration. For the purposes of this response, we refer to "AI Developers" to include AI developers of training models, adapting models, deploying models and each

separate organisation involved in the AI lifecycle. Personal data of BCC creators/performers includes data such as names, likeness, voice, as well as potentially sensitive category data including biometric data.

We believe the Information Commissioner's Office (ICO) has a duty as a regulator to scrutinise the training practices of generative AI application developers and the use of personal data. The BCC is of the view that it is vital that the ICO takes on the responsibility of upholding data protection standards which is extremely important to the creative industry (as well as society at large to protect individual data subjects) particularly in the absence of a dedicated regulator safeguarding the interests of creators.

We agree that there should be a delineation of practices at the ingestion stage, when data including personal data undergoes processing. Compliance would mandate that developers ensure their processing activities are not only in compliance with pertinent laws, including copyright, but also is undertaken under valid lawful basis consistent with the UK data protection framework.

We consequently agree with the ICO's restatement that "as part of complying with the lawfulness principle of data protection, developers need to ensure their processing":

- is not in breach of any laws;
- and has a lawful basis under UK GDPR".

Our position

- **Transparency is vital. How creative works, performances and the personal data associated with these works (which are protected by copyright) can be ingested and adapted in the training of generative AI applications process must be clear and transparent and diligent record keeping must be in place.**
- **This can be accomplished by respecting existing UK copyright and data protection law frameworks. IP licensing safeguards will remain vital to protect against the unfair use and devaluation of copyright protected work.**
- **For clarity, personal data includes the voice, image, likeness of a creator/performer and their creative style if it refers to an identifiable individual (for example 'Splash Painting California' being identifiable as David Hockney) and of course biometric data.**
- **We refer to the human right to privacy, which has been recognised at both the international and national level, including under Article 12 UDHR.**

Consultation questions

a. The right to be informed

Web Scraped Data and Copyright Law: The procurement and processing of personal data for the purposes of AI training, without the knowledge of or securing permission from the person whose data is being scraped, runs counter to the initial step of compliance with the lawfulness principle of data protection. AI training datasets are commonly created by scraping the internet for data. Moreover, data scraping often violates the explicit prohibitions outlined in the terms and conditions of scraped websites.

As we outlined in our previous submissions without express permission to process personal data, there is generally no lawful basis under UK GDPR. AI developers risk failing to be transparent about how and why that might meet a "legitimate interests" threat under UK GDPR rules.

We agree that resource or expense requirements should be factored into AI developers business decisions from the very start and that the default approach must ensure data protection principles – including transparency, fairness and accountability - are implemented effectively from the start including explicit consent sought from rights holders from the outset. It would be perverse if AI developers rely on the alleged disproportionality of the measures required under data protection rules to undermine the privacy rights of individuals.

We agree with the ICO that “(g)enerative AI developers seeking to apply Article 14 exceptions still must take appropriate measures to protect individuals’ rights and freedoms, including by making privacy information publicly available.”

b. Rights to erasure, to rectification, to restriction of processing and to object to processing

As we stressed in our response to the third consultation on accuracy, record keeping of the ingested material by AI developers is an indispensable condition underlining this right. As we noted regarding the right to be informed – providing accurate records is an aspect of operating in the market; respecting data protection and privacy laws in the UK.

Whilst we recognise the “challenges” for AI developers, we suggest that these challenges could be addressed e.g. by providing central records instead of responding to individual requests as general initial activity. An individual response might remain required under the data protection framework. It is worth remembering that up until

now AI developers have been processing data without express consent for data protection and for other purposes (such as copyright and contractual obligations).

Conclusion

- **AI developers are routinely using personal data without express permission whether from an individual or from controllers responsible for the protection of personal data for which they are responsible.**
- **This is against existing UK copyright and data protection law frameworks (in particular in the absence of legitimate interests for the data processing).**
- **This is of particular relevance for creators and performers where personal data includes voice, likeness, biometric data and individual style for example.**
- **There is an urgent need for such creative companies to be able to secure information from data scrapers to show that personal data processing rules are being observed and that unwitting liabilities are not attached to breaches by creative companies over which they have no control.**
- **AI developers as data controller are ultimately responsible for ensuring transparency and diligent record keeping of ingested materials as well as ensuring data protection principles are considered from the outset to protect personal data and respect existing UK copyright and data protection law frameworks.**