

## **British Copyright Council: ICO consultation series on generative AI and data protection**

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.

Many BCC members are creators who increasingly work with AI technologies as both assistive and generative 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 in place of directly licensed human-authored works. As such, transparency over how creative works can be ingested and adapted throughout this process will be increasingly important, and IP licensing safeguards will remain vital to protect against the unfair use and devaluation of copyright protected work. This can be accomplished by respecting existing UK copyright and data protection law frameworks.

We appreciate this timely consultation, considering the recently released Government response the AI Regulation White Paper and the fact that data protection has emerged as a crucial area where generative AI has impacted creators and rights holders.

Creative works are frequently being ingested for training generative AI applications without securing permission for copyright licensing or obtaining consent for personal data processing. Moreover, the extraction of creative works such as text, film, image and music from publicly available websites, even in cases where such practices are explicitly prohibited in their terms and conditions, compounds the issue even further. Such creative works include personal data intrinsic to the individual creator and, depending on nature of the work, sensitive category data. The fact that creative works may be public accessible online for a specific authorised use does not mean it is "publicly available" for scraping.

We believe the Information Commissioner's Office (ICO) is in a strong position as a regulator to scrutinise the training practices of generative AI application developers. This statement is in line with the government's outlined strategy articulated in the AI White Paper, which delineates substantial roles for specific regulators. From our perspective, the ICO assumes the responsibility of upholding data protection standards which is extremely important to the creative industry (as well as society as 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"

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

### **(a) Copyright Law**

The procurement and processing of personal data for the purposes of AI training, without securing permission, 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. This process often encroaches upon numerous rights which, in our opinion, would require explicit consent from rights holders. Moreover, data scraping often violates the explicit prohibitions outlined in the terms and conditions of scraped websites.

### **(b) Lawful Basis under UK GDPR**

Furthermore, AI developers risk failing to meet each limb of the three cumulative tests essential for establishing "legitimate interests" under UK GDPR rules.

#### *Purpose Test: Is There a Valid Interest?*

While data scraping may serve the commercial interests of developers, it fails to align with the interests of data subjects, creators, or rights holders protected by UK GDPR. Notably, there is also a societal need to protect against the misuse of personal data, especially concerning deep fakes. This is even more pertinent given that at the point at which the personal data is scraped, particularly in relation to creative works (as opposed to medical or scientific data for example) the AI developer may have no clear or specific purpose for such processing beyond a vague or generic model.

#### *Necessity Test: Is Web Scraping Necessary Given the Purpose?*

Web scraping, absent of consent, fails the necessity test. Developers, like any other data protection controller or processor, have the agency to seek the requisite form of consent for processing personal data, and it remains within the remit of the data subject to either grant or withhold consent.

#### *Balancing Test: Do Individuals' Rights Override the Interest of the Generative AI Developer?*

Data protection aims to protect the rights of individual data subjects against potential abuses of their personal data and rights such as personality, intellectual property, and contractual freedoms. It is imperative that these rights be preserved.

It is difficult to envisage a scenario where the AI developer's commercial interests in generating AI material – in any of the 3 scenarios outlined – would outweigh the individual data subject's fundamental rights and freedoms and their reasonable expectation as to how their personal data, including sensitive category data could be used and their rights as data subjects could be enforced.

Under the provisions of the UK data protection framework any processing or storage of personal data such as the voice or individual style of creators/ artists requires the express consent of the individual concerned.

We are looking forward to working with the ICO as the relevant regulator protecting the personal as well as biometric data of UK creators and artists.