

News Media Association Response to the Data (Use and Access) Bill Committee’s Call for Evidence: Retaining the Kidron Amendments to Ensure News Publishers, the Broader Creative Industries, and Generative AI Developers Can Innovate and Grow Together

Prior to Lords Report Stage, the Data (Use and Access) Bill (“**the Bill**”) was silent on the most pressing data policy issue of our time: the unlicensed and illegal use of copyrighted content by Generative AI (“**GAI**”) developers. This is intellectual property (“**IP**”) theft on a mass scale. Clauses 95(1), 135, 136, 137, 138 and 139 - voted into the Bill at Lords Report Stage - will make the UK’s existing copyright regime enforceable. **We ask members of the Committee to vote against Government amendments 22, 23, 24, 25, 26 and 27 which would remove these critical provisions from the Bill.** You can find an in-depth description of these amendments in Annex 1.

Key Points

Unlicensed Use of Copyrighted Content: GAI developers are using copyrighted content without licenses, amounting to mass IP theft. [YouGov polling published by the Creative Rights in AI Coalition](#) shows 72% of the UK public believe GAI companies should be required to pay to use the content used to train GAI, while 80% said GAI companies should be required to make public all the information that their models have been trained upon.

Government Consultation: The government proposes a text and data mining (“**TDM**”) exception, allowing GAI firms to use content unless rights holders opt-out. This approach is unevidenced, unworkable, and incredibly regressive. Individual creators would face insurmountable challenges in managing opt-outs. The government’s solution transfers the burden of copyright from the largest tech firms in the world to the smallest creators, leaving no time for creativity itself.

Copyright and AI Provisions in the Data (Use and Access) Bill: Clauses 95(1), 135, 136, 137, 138, and 139 retain and enforce the existing copyright framework, which is clear that GAI firms must seek a license to use copyrighted content. They are based on three key principles:

1. If you want to build a GAI model in the UK, or access the UK market, you should comply with our laws. That means asking creators for permission before you use their immensely valuable content.
2. GAI developers should be transparent about the individual works used to train their models, so that creators can check if their work has been used illegally – this is the only way to ensure that GAI firms comply with the law.
3. Compliance with copyright law, and compliance with provisions designed to enforce it, should have regulatory heft behind them. Without significant repercussions for non-compliance, GAI firms will continue to break the law.

GAI Innovation: High quality data is essential for GAI success, but the supply of public, human-created text data is depleting. If high quality creative content is not paid for, then less and less will be created, and GAI innovation will stall. Licensing is feasible and necessary, contrary to GAI firms' claims – they just don’t want to pay. GAI firms know it would seem ridiculous if they asked

for free electricity, or free labour, or free computer chips, but they've seen an opportunity to fool governments all over the world over into thinking that they shouldn't have to pay for the component of AI that is more essential than any other – the data.

Investment Myths: Weakening copyright law won't attract more GAI investment; other factors like energy costs and access to skills are far more significant. International tech firms will just take the copyrighted content and run. Meanwhile, UK AI firms that are building genuinely innovative tech to make advances in science, medicine, and climate change have no need for copyrighted content. This is a policy written by Big Tech, for Big Tech.

Growth: The government's policy will transfer value to Big Tech firms away from the 2.4 million people in the UK who work in the creative industries. We must ask what kind of growth we want to see – the kind that goes straight to Silicon Valley shareholders, or sustainable growth that benefits creators across the UK? The £126 billion creative industries is at stake, and the government must urgently produce a full economic impact assessment of their plans.

The Government's Copyright and AI Consultation: Degrading Copyright and a Vague Commitment to Transparency

The government's consultation is based on the mistaken idea - promoted by tech lobbyists and echoed in the consultation - that there is a lack of clarity in existing copyright law. This is completely untrue: the use of copyrighted content by GAI firms without a licence is theft on a mass scale, and there is no objective case for a new TDM exception.

The government has stated that ongoing court cases are evidence of uncertainty. But the government does not conclude that existing law is completely obsolete every time an individual or business breaks the law and then chooses to defend themselves in court. Instead, as in other instances, we should make the law enforceable.

Yet the government is proposing to change the UK's copyright framework by creating a TDM exception where rights holders have not expressly reserved their rights - in other words, an 'opt-out' system. But if the government truly believes that there is uncertainty in the existing law, then it would not be proposing a rights reservation. Under a rights reservation, GAI firms will still claim that there is uncertainty in the underlying law and ignore attempts to opt-out. The government should withdraw its proposals if it cannot specify precisely what uncertainty it is trying to address.

No effective rights reservation - or opt-out - system for the use of content by GAI has been proposed or implemented anywhere in the world, making the government proposals entirely speculative. What the government is proposing is an incredibly unfair trade-off, giving the creative industries a vague commitment to transparency whilst giving the rights of millions of creators to GAI firms. Creators are desperate for a solution after years of copyright theft by GAI firms, but making a crime legal cannot be the solution to mass theft.

The Unworkability of an Opt-Out Regime for Text and Data Mining

As well as being entirely unfair, Fairly Trained founder Ed Newton-Rex has set out that a rights reservation (opt-out) system would be [unworkable for several reasons](#), including:

1. **GAI developers have already ignored publishers' attempts to opt-out of scraping without a licence:** Mass illegal scraping has compelled publishers to use web protocols like robots.txt to try and prevent scraping, but these protocols are voluntary, unenforceable, and often ignored.
2. **It is not possible to opt-out downstream copies of copyrighted works:** Even if news publishers could opt-out of the scraping of their content without a licence from their own websites, it is not possible to prevent scraping from pirate websites, or when an article is shared on social media.
3. **Emerging technologies:** Meta has indicated that it will use data captured by its Meta Smart Glasses to train GAI. There is no way to signal an opt-out when data is captured by a camera on smart glasses, which is a rapidly growing market.
4. **The complex web crawler landscape:** There will always be a time lag between the deployment of a new web crawler by a GAI firm and a news publisher opting-out, creating a window of opportunity for GAI developers to scrape without a licence.
5. **Administrative burden, particularly for individual creators:** In February 2024, over 40% of the top 100 English language news websites did not block any GAI crawlers. This lack of awareness will only be exacerbated for smaller news publishers and individuals.

The Copyright and AI Provisions in the Data (Use and Access) Bill: Making the Existing Copyright Framework Workable with Meaningful Transparency

Baroness Kidron's amendments, which were voted into the Data (Use and Access) Bill at Lords Report Stage, would retain the current copyright framework for TDM - where the onus is on GAI firms to seek a licence - and expand it to cover all GAI models marketed in the UK. This is the only way to guarantee that rights holders have control: strong copyright law is the most effective - and the only effective - rights reservation mechanism in existence.

These provisions are complemented by measures that provide meaningful transparency, at a level of granularity that allows a rights holder to understand when their work has been used. The amendments would also provide transparency over the crawlers that scrape content. Robust enforcement provisions will allow rights holders to hold GAI firms to account, incentivising GAI firms to come to the table and negotiate for the IP that is essential for their success.

In short, instead of trading away copyright protections for a vague promise of transparency, Baroness Kidron's amendments promise meaningful transparency provisions that would make the UK's existing gold standard copyright regime enforceable.

UK Competitiveness, GAI Innovation, and Growth

The UK has an opportunity to lead the way, inspiring other nations to adopt high standards and amplifying our influence on the global stage. There is no 'balance' to be struck between creators' copyrights and GAI innovation: IP rights are central to GAI innovation.

The Impotence of High Quality Data to GAI

This is clear, because high quality data is essential to the success of GAI. And this data is running out. Researchers have forecasted that, if current trends persist, GAI developers might deplete the available stock of public, human-created text data between 2026 and 2032. Concurrently, the amount of GAI-generated content on the internet is expected to increase. This could not only hinder innovation in GAI by limiting its growth potential but also degrade the quality of GAI. This degradation, known as ‘model collapse’, happens when GAI models start training on their own, lower quality outputs. This will see GAI innovation stall and stymie the growth that the government desires.

This scenario will become reality if trusted news outlets are not compensated for the use of their content, as they will be unable to keep producing it. Yet even if this were not the case, GAI firms should not demand the use of content for free simply because it may improve their service: music and TV streaming services do not presume to use content for free, simply because it would improve their service. GAI should be no different.

As a result of the mass theft that has taken place, is estimated [that AI firms only spend 0.1% of their budgets on content licencing at present](#). Yet licencing is clearly feasible, as evidenced by the [deals that have taken place, and the growing number of companies created to facilitate licencing](#). This shows that GAI firms’ arguments that licencing is unfeasible are entirely bogus. Illegal scraping is a choice intended to maximise shareholder value, not a necessity.

The Myth of AI Investment

It is also entirely unproven that degrading copyright law will attract more GAI investment into the UK. There are myriad other more significant factors – the cost of energy, access to skills, access to compute, and finance – that will determine if GAI firms invest here. We should not sacrifice the UK’s creative crown jewels on the basis of speculative promises of investment.

Further, there is a clear distinction to be drawn between the very useful (potential) benefits that AI (in a broad sense) can bring, and the need to scrape content for GAI. The copyright debate should be focused on the provision of creative data for consumer-facing products that are designed to replace the very content that they use, rather than products that represent significant innovation.

For example, the benefits that AI may bring in climate change or in health will be brought by very focused, niche AI models using relatively narrow datasets. The data that will build AI to revolutionise science, medicine, and finance is entirely distinct from the desire of Big Tech incumbents to scrape creative, copyrighted data for free to develop mass-market products.

UK AI, the trade body representing the broader British AI industry (i.e. beyond Big Tech incumbents), has voiced strong opposition to the government’s proposals, saying: “The opt-out model would significantly harm the creative sectors to achieve a minimal gain for a handful of global tech companies. This creates unnecessary conflict which will erode public trust in the nascent AI sector, limiting its future growth. In the long term, this will undermine economic growth more than any benefits the opt-out could deliver”.

Matt Clifford's AI Opportunities Action Plan

Matt Clifford's AI Opportunities Action Plan has cast doubt over the status of the government's Copyright and AI Consultation, with Recommendation 24 stating that the UK's TDM regime should be reformed 'so that it is at least as competitive as the EU'. The Plan also restates the misleading view that there is 'uncertainty' in UK law, adding that the EU approach is 'designed to support AI innovation' and that the 'UK is falling behind'.

As [set out by the Creative Rights in AI Coalition](#), a new coalition of over 50 creative organisations: 'The EU is still working out how to implement its EU AI Act and there are persistent questions over the workability of their opt-out regime. This serves as a real-time warning for the government about imitating regimes that have shown no signs of being effective. The UK should learn from the EU regime's shortcomings, not blindly imitate it'.

It is also misleading to say that the UK is falling behind. The EU is still working out how its regime will apply, and the US courts have yet to determine what constitutes 'fair use'. The UK's position, with the second biggest creative industries and third biggest AI sector globally, means we have a unique opportunity to create solutions that drive growth in both the creative and tech sectors.

International Law

IP expert Nicholas Caddick KC has published a legal opinion setting out that the government's plans would conflict with the UK's obligations under the Berne Convention, an international treaty. Caddick states that the opt-out is "clearly prohibited" as - by requiring rightsholders to reserve their rights - the opt-out would impose a formality on those rightsholders' enjoyment and exercise of those rights. Article 5(2) of the Berne Convention does not allow this.

The Emerging Market for GAI Training Data

GAI firms have sought to argue that it would be too complex or costly to license copyrighted works for train their models. The emerging market for GAI training data demonstrates that this is plainly untrue. Some GAI firms have signed significant deals with many news publishers, demonstrating: (a) that training data has a calculable monetary value; and (b) that it is possible to agree licenses for this data.

There is no market failure, but the government will be sweeping the rug out from under creatives in this nascent market if it proceeds with its plans. Myriad creative businesses are on record saying they would be happy to license their work, and a number of companies - such as Human Native and Created by Humans - exist to facilitate deals between creators and GAI firms.

The key factor hampering the further development of the market for high quality training data in the UK is the lack of transparency and robust enforcement mechanisms to enforce our existing copyright law, as well as a mistaken belief that training abroad should exempt a GAI firms from paying for UK content. Clauses 95(1), 135, 136, 137, 138 and 139 would remedy this, and should remain in the Bill.

Annex 1 – An In-Depth Description of Clauses 95(1), 135, 136, 137, 138 and 139

There is no need to change the underlying UK copyright framework to drive a dynamic licensing market for copyrighted content. Instead, six provisions added to the Data (Use and Access) Bill at Lords Report Stage would make the UK's existing, gold standard copyright regime enforceable in the age of generative AI (“**GAI**”). *Whilst these provisions have been voted into the Bill, for ease, the hyperlinks take you to the individual clauses as tabled.*

[Clause 95\(1\) - Analysis of performance](#)

This provision would require a report on the Information and Commissioner's Office (“**ICO**”) regarding its performance of the duties assigned to it by clause 64 (see below).

[Clause 135 - Compliance with UK copyright law by operators of web crawlers and general-purpose AI models](#)

Compliance with UK copyright law

This clause will ensure that the operators of web crawlers (the internet bots that copy news publishers' websites) and general-purpose AI models comply with UK copyright law, including the Copyright, Design and Patents Act 1988 (“**the CDPA**”), regardless of the jurisdiction in which the scraping of content or GAI training takes place. Section 29 of the CPDA is clear that text and data mining (“**TDM**”) – the technique used to copy and extract data for GAI training – is only allowed without a licence/permission for non-commercial research.

The jurisdictional issue

GAI developers have sought to argue that they should not have to comply with UK copyright law if the training of their models takes place in another jurisdiction. Whilst creators strongly dispute this argument, this clause will ensure that GAI developers have to license the data that fuels their models in accordance with UK copyright law if they market their product in the UK, regardless of where the training takes place. This will also ensure that GAI start-ups based in the UK are not put at a competitive disadvantage due to the ability of international firms to conduct training in a different jurisdiction.

Retrieval-augmented generation

This clause will also ensure that UK copyright law applies to the entire lifecycle of a GAI model, including retrieval-augmented generation (“**RAG**”). This is crucial, as news content is immensely valuable in making GAI models more reliable after the initial training. GAI search tools like Perplexity or Google's AI Overviews use RAG to connect an underlying model – such as Google Gemini - with fresh publisher content after a user makes a query in Google Search, for example. A response is then provided by stitching together the content scraped in RAG with data from the model. Scraping news content in real time is essential for creating GAI search services because their underlying models alone are: (a) immediately out of date upon release; and (b) prone to hallucination.

Google places ‘AI Overviews’ above links to publishers' websites in Google Search, using its market power to compete with the very publishers whose data is essential in creating AI

Overviews. Google is also selling RAG data to Meta for use in its AI, further profiting from copyrighted content without remunerating publishers.

Clause 136 - Transparency of crawler identity, purpose, and segmentation

Transparency over crawler identity and purpose

This clause will require operators of internet crawlers and general-purpose AI models to be transparent about the identity and purpose of their crawlers; operate distinct crawlers for different purposes; and not penalise copyright holders who choose to deny scraping for GAI by downranking their content in, or removing their content from, a search engine.

At present, the operators of crawlers that scrape news content and other copyrighted works can obscure their identity, making it difficult and time-consuming for news publishers to identify and combat illegal scraping. This clause will incentivise a dynamic licensing market by allowing copyright holders to easily identify illegal scraping and seek redress.

Crucially, the clause will also ensure that the operators of crawlers disclose the businesses to which they sell the data they have scraped, making it more difficult for GAI developers who purchase illegally scraped content to avoid compliance with UK copyright law. It will also give copyright holders a point of contact to complain about scraping without a licence, increasing accountability.

Crawler segmentation, and preventing discrimination

The clause will also ensure that the operators of web crawlers and general-purpose AI models deploy distinct crawlers for different purposes, and prevent the operator of a search engine from removing or downranking content from a news publisher that has chosen not to allow their content to be scraped for GAI training.

This is crucial, as Big Tech firms leverage their market power to scrape the web for GAI training without a licence. The same crawler that: (a) indexes news content in ‘traditional’ Google Search results; also (b) scrapes publisher data for Google’s ‘AI Overviews’ (GAI generated news summaries that are placed at the top of Google Search, above the links to news articles which Google scrapes to create their AI answers). Google does not allow publishers to: (a) prevent scraping of their content AI Overviews; whilst (b) still allowing their content to be indexed in Search results. This effectively forces publishers to allow their data to be used in AI Overviews, which are intended to be a substitute for their high quality journalism.

Clause 137 - Transparency of copyrighted works scraped

If GAI developers use copyrighted content without permission, publishers must be able to identify this illegal use and seek redress. Meaningful transparency must provide a level of granularity far beyond existing data cards or model cards currently provided by GAI firms. This clause will be crucial in making UK copyright law enforceable by requiring operators of web crawlers and general-purpose AI models to be transparent about the copyrighted works they have scraped.

Specifically, this clause would ensure GAI firms are compelled to share the individual URLs they have accessed to scrape copyrighted content, and further information “that can be used to

identify individual works”, as well as the timeframe of data collection. Such transparency will ensure that GAI developers can be held accountable for copyright theft: this is the key element in incentivising a dynamic licensing market.

This mechanism could work like the Subject Access Request (“**SAR**”) system used by the ICO, with rights holders first approaching GAI firms for information and only going to the regulator if the firm refuses to comply. Such a system would counter claims by GAI firms that providing transparency would compromise trade secrets or create safety and security concerns. The SAR system is a good example because it doesn't require legal expertise from the requester, which will be crucial for individual creators seeking to understand when their work is used illegally.

Clause 138 - Enforcement

Enforcement of these provisions – both via a regulator as well as through private enforcement by copyright owners - will be critical to their success.

This clause applies powers already exercised by the ICO under the Data Protection Act 2018 to the above provisions, with significant penalties – including fines – for non-compliance. The use of enforcement notices will allow the ICO to require GAI firm to take steps to remedy instances of non-compliance. In addition, this clause allows copyright holders to take private action against a GAI firm that fails to comply with the above provisions, to recover damages, and to secure injunctive relief.

This will complement regulatory enforcement, further incentivising GAI developers to comply with the law. Absent robust enforcement mechanisms, GAI developers will not be sufficiently incentivised to comply with UK copyright law.

Clause 139: Technical solutions

This clause requires the Secretary of State to conduct a review of technical solutions, such as watermarking, that could allow copyright holders to identify unauthorised scraping by GAI firms. Such technical solutions could aid copyright holders in asserting their rights.

For clarity, under UK copyright law as it stands – or as expanded to all GAI firms marketing their models in the UK – it would not be necessary for copyright holders to utilise these technical mechanisms to reserve their rights. However, such solutions could complement, and aid the efficacy of, the transparency provisions set out above, giving copyright holders greater clarity over how their works have been used.