Copyright and AI Mythbuster

The Creative Rights in AI Coalition is a broad group of rightsholders, AI businesses, and organisations from across the creative industries. We believe the Government can spur growth in the creative and tech sectors by maintaining copyright.

The copyright and AI debate is often framed as 'creator vs machine', or 'head vs heart'. On the one hand, most are sympathetic to the idea that creators should have control over their work and be able to earn a living from it. On the other, many believe that these interests must be 'balanced' with the need to support AI innovation and AI adoption in the UK.

In fact, far from being a barrier to AI growth, strong creative copyright law is central to generative AI innovation (and is not at all relevant to the many other forms of AI that the government sees as central to its economic growth mission). Ultimately, we do not need to 'compromise' with an unworkable and regressive opt-out model that will only benefit a handful of already dominant Big Tech firms. We can support British creativity and British AI without compromising creative rights.

Myth 1 - AI training is just like a human learning from creative work, and the creative sector will simply evolve in response to AI

Comparing AI training to human learning – for example, someone gaining knowledge through reading a library book – is inaccurate in two key respects. First, the mass scraping of millions of works is on an entirely different scale to a human reading a book. Second, the books in a library, or the paintings in a gallery, have been paid for.

It has been claimed that it is unnecessary to remunerate creators as they are already using AI in their work, and the creative sector will simply evolve. Yet just because creators are AI adopters does not mean that they should be content to rent back their own work through AI models that have used their creativity for free.

It is true that creators have adapted to advances in technology, such as the diffusion of photography; but a camera is not made from stolen paintings, it is an artist's expressive tool much like a paintbrush or stylo, and photography has been acknowledged as an art form for some time. In contrast, generative AI models do not 'create', they are prediction engines which rely on creative inputs to manufacture creative outputs.

This means that the unlicensed use of creative content by generative AI firms - which then seek to substitute for the very content they have ingested - puts creative livelihoods at risk. In February 2025, 58.1% of Association of Photographers members reported losing work as a direct result of generative AI, with an average value transfer (loss) of £14,400 per photographer. Of course, it is not unreasonable for businesses to utilise generative AI tools instead of human creators, but the creators of original material used for model training and model grounding must be remunerated for the value that their work brings to AI models.

Myth 2 - There is a lack of clarity in existing copyright law that is preventing creators from asserting their rights

UK copyright law is completely clear that commercial text and data mining – the technique used to gather and analyse data for AI training – requires permission from the rightsholder. There is an

exception for non-commercial research use, facilitating scientific research, but this obviously does not apply to Big Tech-owned and backed models with multi-billion-dollar valuations.

There is also an exception for 'temporary copying', intended to allow 'transient or incidental' copying which has no independent economic significance, and is intended to enable the 'lawful use of the work'. This exception was introduced to enable web browsing, allowing users to view webpages. It clearly does not apply to generative AI models that are intentionally trained on web-scraped data that is not deleted after training, and which have immense economic significance.

The key barrier to creators asserting their rights is a lack of transparency over the individual works that have been used to train AI models. If this information were available, it would be far simpler for courts to hear cases and apply UK copyright law.

Myth 3 - It is not feasible for AI firms to agree licenses for the content they use

AI firms often claim that it is simply not feasible to license all of the copyright works in an AI model, yet in most cases, they have not tried to license content before using it. Also, many AI firms have already agreed licenses with news publishers and others for the use of their content in the UK (which they would be unlikely to do if they truly considered UK copyright law allows the use of content without a license).

There is a clear willingness on the part of rightsholders to license content – they are not opposed to AI training but are opposed to AI training without consent or remuneration. In a poll of 1,000 artists by the Design and Artists Copyright Licensing Society (DACS), 84% said they would be willing to sign up to a license for their work to be used in AI training.

Publishers' Licensing Society (PLS) and the Authors' Licensing and Collecting Society (ALCS) have recently announced the development of a collective license for generative AI training, following a poll which showed 81% of ALCS members would want to be part of a collective licensing solution. The development of efficient and effective market-based solutions by collective licensing organisations, as well as the growth in firms acting as a marketplace for training data – such as Human Native and Defined.ai – dispels myths about licensing being unfeasible or overly complex.

Myth 4 – Individual works have no real value in AI training

AI firms have claimed that the value of individual works in AI training datasets is nugatory, with the value coming from the overall size of the dataset. This is akin to arguing that individual contributors to a magazine or a single violinist in an orchestra should not be paid, simply because their contribution is combined with other people's work. In fact, one forecast predicts that the global AI training dataset market will reach $34 billion by 2033 (and this is without the development of policies to enforce robust copyright law).

This argument also ignores key stages in model training such as fine-tuning (where a pre-trained model is trained on smaller, high-quality dataset to adapt it for a specific function) or retrieval-augmented generation (known as RAG, where a model retrieves information in real time in response to a query, improving accuracy and allowing it to provide up-to-date information). High-quality data is particularly valuable in these training stages.

Beyond the question of value, there is the even more fundamental question of control. It would be contradictory to demand that a rightsholder relinquish control over their work because an AI firm claims that unlicensed scraping is integral to AI innovation, whilst simultaneously dismissing the suggestion that a creator's work holds any value. Instead, we should create the right conditions to let the market decide the value of creative content.

The UK's goldmine of creative content – in particular, our English language content – is incredibly valuable to AI firms. If using creative content really is so crucial to the functioning of AI models, then it stands to reason that AI firms will ultimately be willing to pay for it. The UK can be an AI superpower, but not just as a host for, and deployer of AI models developed abroad: we can hold a pivotal position as the provider of creative content for generative AI models around the globe.

Myth 5 - An opt-out will give creators control

The Government freely admits that an effective opt-out mechanism – which would allow creators to easily and effectively prevent the use of their work without a license – does not currently exist. The EU has had an opt-out in place since 2019, yet no effective mechanisms have developed in that time. It is odd that the government has chosen a preferred policy option, which is contingent on developing technologies that are by no means certain to come into existence.

In any case, an opt-out is incredibly regressive, putting the burden on the smallest creators to control the use of their work by the most powerful tech firms in the world. It is akin to expecting a record shop owner to take steps to enforce the Theft Act to prevent LPs being stolen, rather than stealing simply being illegal. There will be immense time pressure to make a decision before training takes place (if indeed any pre-warning is given), and the administrative burden will be immense. Such burdensome regulation will leave less time for creativity, harming creator productivity.

Even if a creator can opt-out content on their owned and operated website with certainty, the downstream copies of creative work – on a pirate website, or when posted on social media – will remain vulnerable to unlicensed scraping. The LibGen database of millions of books and academic papers, used by Meta in its AI training, is a key example. And social media sites such as Instagram are using posted content for AI training, with no opportunity to opt-out.

Robots.txt, the most commonly used opt-out tool which signals that certain URLs should not be used for training, is routinely evaded or ignored. And it gives rightsholders no control over downstream uses of their work. There is also a danger that emerging technologies will make location-based opt-out mechanisms like robots.txt, or unit-based-opt out mechanisms like adding metadata to a photograph, will swiftly become obsolete as technology advances. Meta has indicated that it will use data captured by its Smart Glasses in AI training, and it will be impossible for these glasses to capture robots.txt or metadata opt-outs. Furthermore, AI-developers can simply ignore - and routinely have ignored - machine-readable protocols without legal enforcement measures to uphold rightsholders requests.

In short, the myriad barriers to an effective opt-out make it unlikely that any mechanism will soon develop that meets the Government's stated aims of "effective, accessible, and widely adopted". The Government has stated that it will not introduce an opt-out without these standards being

met; given the severe doubts around opt-outs, alternative options – such as enforcing existing copyright law with transparency provisions – must be considered with urgency.

Myth 6 - We must mirror more permissive copyright regimes in other jurisdictions

It is argued that the US fair use exception, and weaker copyright law in other jurisdictions, mean that the UK must move to a more permissive copyright regime. Yet recent US developments such as Thomson Reuters vs ROSS Intelligence decision show it is highly likely that the court cases around 'fair use' will lead to the US effectively having a similar copyright law with regards to AI as the UK. American publishing representatives have also noted that the previous Donald Trump administration was strong on IP protections for rightsholders.

Whilst the EU has an opt-out, it is far from clear how this will apply in practice, as the EU AI Act provisions on transparency and opt-out mechanism have yet to come into force. Axel Voss, an MEP involved in the creation of the 2019 Copyright Directive, has said the opt-out was a "misunderstanding" that was intended to have a limited private use. This contradicts Matt Clifford's AI Opportunities Action Plan, which spuriously claims that the opt-out was created to spur innovation, despite it being created before the advent of generative AI. The EU is now desperately – and so far, unsuccessfully – trying to make the opt-out workable through the EU AI Act.

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. A key element of this is extraterritoriality: the idea that copyright law should apply to any AI model deployed or marketed in the UK, regardless of where the training takes place.

Of course, there will always be AI firms that do not market their models in the UK and will never fall under the scope of UK law; however, this is no reason not to apply the law to the many frontier models that are developed by firms desperate to access the UK market. Respectable businesses with international reputations will still want to trade in lucrative markets, even with robust copyright law and meaningful transparency requirements.

Myth 7 - Our national security will be put in danger if we retain and enforce strong copyright law

Concerns have been raised about the implications for our national security if advanced AI models cannot access copyright works for free. Yet models concerned with key areas such as security, defence, science, health, and public services simply do not require creative content to function.

For example, advanced medical research does not need creative copyright, far less creative content scraped indiscriminately from the web. AlphaFold, often cited as a key UK AI breakthrough, uses meticulously selected and legally sourced data, such as the Protein Data Bank. Models that will improve public services will have little need for creative copyright. They will use data held by the government, such as NHS medical records - a taxpayer-owned asset that could provide vital new resources back into the NHS in exchange for access to data. Moreover, UK government agencies concerned with national security issues will be building their own secure models.

The real threat to our national security would be becoming completely reliant on a handful of US Big Tech models at a time of such heightened geopolitical tension, when it is clear the US administration will use any lever necessary to exert influence over other countries. This is compounded by the risks to our information ecosystem if trusted news publishers are not remunerated for their work and produce less quality journalism, and AI models will be more likely to produce misinformation if they are not sustained with up-to-date quality content (which requires them to fund it). This represents a substantial threat to the quality of public discourse (to note, under the EU Digital Services Act, providers of 'very large online platforms' and 'very large search engines are required to conduct an assessment of systemic risks stemming from their systems, including threats to the freedom and pluralism of the media).

Myth 8 - Degrading copyright law will support AI investment

There is no dispute that AI represents a great economic opportunity, but it is simply untrue to state that its success is contingent on weaker copyright law. The key beneficiaries of a sweeping exception to copyright law will be US Big Tech firms. But these firms will not base investment decisions on copyright law. It is energy costs (UK prices are approximately four times higher than the US) and data protection law (much more permissive in the US) that will be the greater determining factors. UK AI - the trade body representing the broader British AI industry - has voiced strong opposition to the government's opt-out proposals, saying they would only bolster Big Tech firms.

Big Tech firms have a history of extracting significant profits from the UK markets, which then flow back to shareholders, rather than supporting long-term, sustainable growth. Promised investment in data centres will lead to very few permanent staff being employed, whilst the construction will likely be undertaken by a small global pool of specialist workers.

Myth 9 - Transparency must be accompanied by an opt-out

There is simply no sound technical or policy reason why transparency provisions – allowing creatives to understand when their individual works have been used – must be accompanied by an opt-out. Transparency would spur a dynamic licensing market for creative content without any need to change underlying copyright law.

As set out above, it is a lack of transparency – rather than a lack of transparency in law – that is preventing legal precedent from developing in the UK. Meaningful transparency would allow our courts to deal with civil enforcement cases swiftly, and very soon the legal risk of infringement would be too great for AI firms to attempt it, meaning smaller creators would not have to assert their rights through the courts. This demonstrates that transparency is not inextricably linked to the opt-out – it could be implemented to enforce our existing laws and prevent the massive harm that is already being done to our creative industries.

There is no technical barrier to transparency, in the same way that there is to an effective opt-out mechanism. The reason the EU has not been able to confirm a Code of Practice on transparency is because AI firms are lobbying frantically to prevent meaningful transparency provisions being created, precisely because it would make them accountable for mass copyright infringement. Given there is no workable opt-out mechanism on the horizon, it would seem entirely sensible to proceed with a transparency-first approach.

Myth 10 - Granular transparency is not feasible, or would compromise trade secrets or model security

AI firms claim that granular transparency mechanisms – allowing individual works to be identified – would be overly burdensome, or could even compromise trade secrets or model security. Yet this information would not need to be disclosed publicly, or in one single data dump. Instead, rightsholders would simply like to be able to ascertain – either through automatic notification by an AI firm, or by submitting a request to an AI firm – which of their own individual works have been used in model training. This can be achieved through private disclosure obligations, which would mitigate any perceived risk from developers of transparency undermining commercial competitiveness or somehow giving away trade secrets.

Moreover, general-purpose AI models are made up of billions of data points: it would be impossible for these disclosures (which would not include the weighting of data, or other key information) to serve any other purpose than incentivising AI firms to comply with the law and ensuring that rightsholders can identify illegal use of their work.

Furthermore, the question of transparency of AI inputs (and indeed the workings of these models through the AI model life cycle) is central to many AI policy areas. The challenges in online safety; data protection; the prevention of bias hiring decisions; stopping AI models from discriminating against people; detecting the presence of harmful or illegal content in training data – all of these policy issues, and many more, can only be tackled if there is meaningful transparency. This government must tackle these issues head on.

Creative Rights in AI Coalition
May 2025
info@creativerightsinai.co.uk