Background card

The UK government's consultation

Currently, developers are subject to copyright law when using large data sets to train artificial intelligence (AI) models. The government has argued how copyright law should apply to training AI is "disputed" and the current rules do not meet the needs of either the UK's creative industries or the AI sector. It has said this uncertainty is limiting investment and innovation in this sector and has made it difficult for rights holders to control how their works are used.

Currently, only a narrow exception to copyright law exists for non-commercial research.

In December 2024, the UK government published proposals to change the way in which this material could be used. This included the establishment of a copyright exemption for AI developers and a new rights reservation model whereby copyright holders would need to opt-out from having their material used for training AI.

This approach would be similar to that adopted by the EU where there is an exception for text and data mining but rightsholders are able to expressly reserve their rights.

An amendment proposed by the House of Lords would explicitly subject AI companies to UK copyright law, regardless of where they are based, reveal the names and owners of web crawlers that currently operate anonymously and allow copyright owners to know when, where and how their work is used.

Main source:
https://lordslibrary.parliament.uk/copyright-and-artificial-intelligence-impact-on-creative-industries/

Characters

POSITION A. IT'S ABOUT THE ECONOMICS

1. **Sir Chris Bryant MP**

About me: I used to be a priest. I have been a Labour MP since 2001. I am now the Minister of State for Media, Tourism and Creative Industries. It was me who made the statement to the House of Commons in mid-December 2024, announcing this consultation. As part of that, I read out lines from a non-existent song by Adele, the Grammy-award winning British singer. I got AI to write the song.

My view: I quoted the Adele lyrics to draw attention to how machines can "imagine" versions of existing artists' songs without paying them any money. I think that our proposed opt-out system will give improved access to content by AI developers, whilst allowing rights holders to control how their content is used for AI training. We intend it to lead to more licensing of content, which is potentially a new revenue stream for creators.

A great outcome for everybody would be a new system of remuneration. This could come from developing a system of simple digital fingerprinting so that people could say 'No, you can't use my work' or 'Yes you can use my work for large language model training once you've remunerated me'. Ideally, creatives could do that individually or collectively, say through DACS (the Design and Artists Copyright Society) or through a musician's record label.

2. **Bertin Martens**

About me: I am a Senior fellow at Bruegel, which is a Europe-wide economic think tank. Until 2022, I was senior economist at the Joint Research Centre of the European Commission. I work on digital economy issues, including e-commerce, digital copyright and media, online platforms and data markets and regulation.

My view: The right to opt-out amounts to economically inefficient overprotection of copyright. Free
use of media content for GenAI training does not affect media sales to consumers. Opt-outs only strengthen the bargaining position of copyright holders, who decide depending on their private interests. That generates windfall profits without any increase in consumer surplus or social welfare.

The licensing of training inputs reduces the quantity of data and the quality of GenAI models, creates transaction costs and reduces competition between GenAI firms. This slows down GenAI-induced innovation in media products and production processes, and productivity gains in all service sectors that apply GenAI. Ultimately, it slows down

economic growth compared to what it could be with competitive and high-quality GenAI.

POSITION B. CULTURAL IMPACT

3. **Kate Mosse**

About me: I am a British novelist, non-fiction and short story writer and broadcaster. I am best known for my 2005 novel *Labyrinth*, which has been translated into more than 37 languages. In 1996 I co-founded the annual award for best UK-published English-language novel by a woman that is now known as the Women's Prize for Fiction.

My view: Five of my novels, including Labyrinth, that I spent 15 years researching, planning, writing, rewriting, editing and publishing, have been illegally scraped to help train large language models.

What is being proposed is like a thief in a corner shop who steals all the Mars bars. When confronted by the shopkeeper, the thief says: “But you didn’t tell me you didn’t want me to steal your Mars bars.”

That’s essentially what the opt-out is. What is being proposed is that creators will have to spend their time hunting down AI companies to see if their work has been stolen. It will take time away from all of us doing the job that we do. It will mean that individuals are less likely to be able to make a living out of their craft. And consequently, everything will simply be diluted. It will be a copy of a copy of a copy of a copy. We will have very little original work.

4. **Ed Newton-Rex**

About me: In 2010 I founded Jukedeck, an AI music generation company that provided music for video, TV, radio, podcasts and games. Most recently I was VP of Audio at Stability AI. I resigned in 2023 over the firm’s belief that taking copyrighted content to train AI models without a licence constitutes “fair use”.

I’m the founder of Fairly Trained, a non-profit that certifies generative AI companies for training data practices that respect creators’ rights.

I am a composer: I write choral music.

My view: I organised a statement from the creative industries warning artificial intelligence companies that unlicensed use of their work is a “major, unjust threat” to artists’ livelihoods. There were 10,500 signatories, including Abba’s Björn Ulvaeus, the actor Julianne Moore and the Radiohead singer Thom Yorke.

There are three key resources that generative AI companies need to build AI models: people, computer power, and data. They spend vast sums on the first two – sometimes a million dollars per engineer, and up to a billion dollars per model. But they expect to take the third –

training data – for free. When AI companies call this 'training data', they dehumanise it. What we're talking about is people's work – their writing, their art, their music.POSITION C. TECH OPTIMISTS

5. **Andrew Ng**

About me: I am a British-American computer scientist and technology entrepreneur focusing on (AI). I was a cofounder and head of Google Brain. In the field of online education, I co-founded Coursera and DeepLearning.AI.

My view: Copyright laws in the United States and elsewhere don't explicitly forbid use of copyrighted works to train machine learning systems. However, the technology's growing ability to produce creative works, and do so in the styles of specific artists and writers, has focused attention on such use and raised legitimate questions about whether it's fair.

The latest advances in machine learning have depended on free access to large quantities of data, much of it scraped from the open internet. Not being able to do this would put the brakes on progress. This would degrade AI's current and future benefits in areas such as art, education, drug development, and manufacturing to name a few.

I believe humanity is better off with permissive sharing of information. If a person can freely access and learn from information on the internet, I'd like to see AI systems allowed to do the same, and I believe this will benefit society. Training AI models should be considered fair use that does not require a license.

6. **Marc Andreessen**

About me: I am cofounder of the venture capital firm Andreessen Horowitz. I co-created the highly influential Mosaic internet browser, one of the first to be widely used internationally. Mosaic was later renamed Netscape. I'm an American.

My view: In 2023 I published the Techno-optimist Manifesto. Here's part of it: "Our civilization was built on technology. Our civilization *is* built on technology. Technology is the glory of human ambition and achievement, the spearhead of progress, and the realization of our potential.

Techno-Optimists believe that societies, like sharks, grow or die. We believe growth is progress – leading to vitality, expansion of life, increasing knowledge, higher well-being. We agree with Paul Collier when he says, "Economic growth is not a cure-all, but lack of growth is a kill-all." We believe everything good is downstream of growth. We believe not growing is stagnation, which leads to zero-sum thinking, internal fighting, degradation, collapse, and ultimately death. There are only three sources of growth: population growth, natural resource utilization, and technology. Developed societies are depopulating all over the

world, across cultures – the total human population may already be shrinking. Natural resource utilization has sharp limits, both real and political. And so the only perpetual source of growth is technology."

The three positions

Position 1: It's about the economics

This view sees AI training as a matter of market economics—copyrighted works are valuable assets, and access to them for training AI should be governed by clear licensing and negotiated deals rather than a strict ban or an unregulated free-for-all.

Key Points:

- Copyright should be treated as an asset for which creators deserve fair market compensation.
- A regulated, contractual framework can allow AI developers to obtain the data they need while creating new revenue streams for creators (e.g. "royalty-sharing AI datasets").
- Government and industry should work together to establish pricing models and licensing standards instead of resorting to an "all or nothing" approach.

Characters:

1. Sir Chris Bryant
2. Bertin Martins

Position 2: It's about the impact on creativity and culture

This view is rooted in the belief that human creativity and cultural integrity are at risk. It holds that AI training on copyrighted works—without explicit, prior consent—undermines the very fabric of creative expression and threatens the livelihoods and legacy of human artists.

Key Points:

- Creators' rights and artistic integrity must be preserved.
- AI companies should not be allowed to "free-ride" on the labour, inspiration, and personal expression of human artists.
- An opt-in system (or other measures ensuring explicit permission) is essential so that the cultural and economic value of human creativity is not eroded by automated systems.

Characters:

3. Kate Mosse

4. Ed Newton-Rex

Position 3: It's about the value of AI: The Tech-Optimist view

This view regards AI as a transformative, public good—akin to a vast digital library—that can democratise access to knowledge and spur innovation. It sees broad access to training data as essential for progress, contending that restrictions on AI training are counterproductive to the advancement of society at large.

Key Points:

- AI training is seen as a form of collective learning rather than simple copying; its outputs can be transformative and serve the public interest.
- Restrictions on data access may hinder innovation and the overall advancement of technology that benefits society.
- The debate should centre on who controls AI—ensuring it remains a public resource rather than being monopolized by a few powerful companies—rather than on narrowly defined copyright issues.

Characters:

1. Andrew Ng
2. Marc Andreesson

Information cards

1. What is generative artificial intelligence and why is it important?

AI enables machines to perform tasks and solve problems that previously required human intelligence. AI applications can curate news feeds on social media, beat any human being at chess, support healthcare workers in analysing x-rays, enable self-driving cars and much more.

Generative AI gathered worldwide attention in 2022 with the release of OpenAI's ChatGPT, as well as text-to-image generators like Dall-E. *Generative* AI takes human prompts and uses it to *generate* a new output, such as an image, video, audio, or written text. See next card for an example.

Generative AI uses huge amounts of training data. ChatGPT trained on over 300 billion words of text data. It identified patterns in the data which it used to generate the output.

Generative AI systems can perform a huge number of different functions: code new programmes, create fake videos, support scientific research, generate art and much more.

Generative AI could transform sectors such as healthcare, climate change mitigation and the arts. However, there are serious risks, including misinformation, bias, automation of jobs, and, long-term, potentially even existential risk.

2. How AI models train on and use large data sets – two views

<u>What the Authors Guild says</u>:

Where AI companies like to say that their machines simply "read" the texts that they are trained on, this is inaccurate anthropomorphizing. Rather, they copy the texts into the software itself, and then they reproduce them again and again.

Also, the bulk of the books used in the "training" datasets originate from pirate sources and websites.

Maya Shanbhag Lang, president of the Authors Guild, said, "It's only fair that authors be compensated for having 'fed' AI and continuing to inform its evolution. Our work cannot be used without consent, credit, and compensation. All three are a must."

<u>What Andrew Ng (Character 5) says</u>:

Today's Large Language Models (LLMs) are trained on a lot of copyrighted text. But copyright law as it relates to generative AI is a mess! Legal challenges are on the rise.

I believe it would be best for society if training AI models were considered fair use (see also 'Fair use' card) that did not require a license. Just as humans are allowed to read articles posted online, learn from them, and then use what they learn to write brand-new articles, I would like to see computers allowed to do so, too.

3. Copyright

Oral societies, such as medieval Europe, tend to view knowledge as the product and expression of the collective, rather than to see it as individual property. However, with copyright laws, intellectual production comes to be seen as a product of an individual, with attendant rights. The most significant point is that patent and copyright laws support the expansion of the range of creative human activities that can be commodified.

A copyright is a type of intellectual property that gives its owner the exclusive legal right to copy, distribute, adapt, display, and perform a creative work, usually for a limited time. The creative work may be in a literary, artistic, educational, or musical form. Copyright is intended to protect the original expression of an idea in the form of a creative work, but not the idea itself. A copyright is subject to limitations based on public interest considerations, such as the fair use (see also 'Fair use' card) doctrine in the United States and fair dealings doctrine in the United Kingdom.

'Fair dealings' covers cases where the copyright infringement is for the purposes of non-commercial research or study, criticism or review, or for the reporting of current events.

4. Fair use in the USA

The fair use principle generally permits use of copyrighted works without permission as long as the use constitutes a significant change in the work and does not threaten the copyright holder's interests.

AI companies tend to make fair use arguments to justify their use of training data. A US Supreme Court decision in Warhol v. Goldsmith casts doubt on these arguments. It rejected fair use as a valid defence in case where the use results in a commercial substitute for the original work.

Here are some examples of AI providing a commercial substitute:

- Traffic to Stack Overflow declined by about 12% after the release of ChatGPT. The average account age of a question-asker trended up after the release of ChatGPT, suggesting that newer members are either not joining or are leaving the community.
- The homework-help and online tutoring website Chegg, had its shares drop 40% after reporting that ChatGPT was hurting its growth.·
- In China, numerous game development enterprises are turning to AI, thereby diminishing the demand for human artists. Within mere months, artist fees plummeted from RMB 20,000 (USD 2,780) per piece to RMB 4,000 (USD 556).

Source for first two examples: https://suchir.net/fair_use.html

5. Examples of licensing arrangements

OpenAI agreed to pay Associated Press for news articles to train its algorithms — an arrangement heralded as the first of its kind. OpenAI will have access to articles produced since 1985, and Associated Press will receive licensing fees and access to OpenAI technology.

In a separate deal, OpenAI extended an earlier agreement with Shutterstock that allows it to train on the stock media licensor's images, videos, and music for six years. In return, Shutterstock will continue to offer OpenAI's text-to-image generation/editing models to its customers.

Some publishers, such as Rupert Murdoch's News Corporation and the Financial Times, have already struck licensing deals to allow Open AI to train its large language models on their journalism.

6. Can AI companies avoid restrictions?

Copyright laws apply where the copying (or training) occurs. Can't AI companies move their operations to jurisdictions with more permissive or unclear copyright rules? Japan for example allows machine learning developers to use copyrighted works whether or not the trained model would be used commercially and regardless of its intended purpose. (This is similar to how companies sometimes base their headquarters in countries with favourable tax laws.)

Since AI models and data can be moved across borders digitally, enforcement becomes challenging unless there is global coordination. This would be similar to how the Berne Convention sets international copyright standards.

In 2023 member states of the Group of Seven (G7), an informal bloc of industrialised democratic governments that includes Japan, announced a plan to craft mutually compatible regulations and standards for generative AI. (That is unlikely to survive the arrival of President Trump of course.)

Other approaches:

1. Countries with stronger protections could refuse to trade with or recognize AI models trained in non-compliant jurisdictions, much like data protection laws under GDPR.
2. If AI-generated content is found to be infringing, copyright holders may still have legal recourse against companies distributing it, even if the training happened elsewhere.

7. **A radical view of copyright**

“Copyright law is a system that worked beautifully for centuries, until AI came along. Copyright law looks like a castle built to protect books, stone walls, drawbridges, guards at every gate. Then AI showed up, not with battering rams, but with wings. And suddenly our castle walls became irrelevant. We don't need better copyright laws – we need an entirely new system.”

The problem

- LLMs use fair use arguments to justify training on copyrighted content. But traditional copyright enforcement can't keep pace with AI's ability to transform content.
- Tech companies' have a "steal first, lawyer up later" approach to data. The discovery phase of recent lawsuits reveals internal discussions about copyright. Companies knew they were using copyrighted material but proceeded anyway
- The "Sue and Hope" model isn't working for creators or platforms
- Small creators are being left behind while tech giants battle in courtrooms

Possible solutions

- Attribution and recognition systems built into AI platforms
- New monetisation models that compensate creators for AI training data
- Blockchain-based content tracking and compensation systems
- Community-driven content models

Source: https://www.theaioptimist.com/p/truth-over-trend-the-year-ai-proved

And here’s a counterview, from the International Center for Law & Economics (ICLE): “Copyright law retains flexibility to adapt to new technologies, as past reforms reacting to photography, sound recordings, software, and the internet all demonstrate.”

8. Arguments against allowing AI models to be trained on any copyrighted material

1. It would undermine the principle that creators have ownership over their intellectual property.
2. Creators would lose control over monetisation. While some may continue creating as a hobby, the professional creative industry could collapse.
3. Creators should have the right to decide how their work is used. Many artists and writers do not want their work used to create AI-generated derivatives, especially if those derivatives are of poor quality, misrepresent their views, or are used in unethical ways (e.g., deepfakes).
4. If AI companies can use vast amounts of copyrighted material for free while human creators must spend years honing their craft, AI gains an unfair advantage.
5. If AI models primarily generate content by recombining existing copyrighted works, there's a risk of homogenization—art becoming a recycled pastiche of what already exists, rather than something truly new. Artistic movements have historically thrived on pushing boundaries, not just remixing previous content.
6. If human creativity is devalued to the point where it's no longer a viable profession, we risk losing the depth, nuance, and emotional authenticity that comes from human-created art.

In conclusion, if AI models can ignore copyright, we risk transforming creativity into a corporate-driven, automated process that serves commercial interests rather than human expression.

Source: ChatGPT (!)

9. Possible common ground between the tech and creative sectors

AI developers want unrestricted access to as much data as possible but creative content owners want just the opposite, as protecting royalty income is critical for them. Where might there be common ground?

First, the two sectors agree on the need for transparency**.**

Second, Chris Bryant told MPs: "If we were to adopt a too tight a regime based on proactive, explicit permission, the danger is that international developers would continue to train their models using UK content accessed overseas, but may not be able to deploy them in the UK … this could significantly disadvantage sectors across our economy, including the creative industries, and sweep the rug from underneath British AI developers."

A different view on where common ground lies comes from a Statement by the Creative Rights in AI Coalition (CRAC): “Retaining the UK’s gold standard copyright protections will create incentives for AI developers to enter into licence agreements with rights holders, ensuring a steady flow of quality, human-authored works for AI training. [The ‘AI optimist’ referred to “The emerging "mad cow disease" problem of AI training on AI-generated content”.] Without this, AI innovation will inevitably stall, and value will drain from both the tech and creative industries which contribute so much to the UK economy and quality of life.”

10. Ways suggested for tackling the problem

The Authors Guild proposes creating a collective license. A collective management organization (CMO) would license out rights on behalf of authors, negotiate meaningful, fair fees with the AI companies, and then distribute those payments to authors. These licenses would cover past uses of books, articles, and other works in AI systems, as well as future uses. The latter would not be licensed without a specific opt-in from the author or other rightsholder.

The House of Lords Communications and Digital Committee said the government should also work with licensing agencies and data repository owners to “create expanded, high quality data sources at the scales needed for LLM training”.

Campaigners for creatives fear a mechanism to reserve, license and be paid for the use of their work in AI training, would probably only benefit the largest rights holders leaving small and medium-scale creators exposed.

Dan Conway, the chief executive of the Publishers Association, said: “There has been no objective case made for a new copyright exception, nor has a water-tight rights-reservation process been outlined anywhere around the globe”.

11. What happens in China?

Since 2018, China has created the Beijing Internet Court. Think small claims court meets digital age. Anyone – individual creator or small company – can submit an AI copyright case online.

Submit your case online, show your creative process, and get a decision that looks at what you actually did – not what some 100-year-old law says about human authorship.

No million-dollar legal teams required. No years-long waiting game. Just practical decisions about real creative work.

Contrast that with the EU’s approach. Aafke Romeijn, a Dutch-language electropop artist, who is on the board of the European Composer and Songwriter Alliance, said she had been told by senior EU officials to take tech companies to court to preserve her copyright. “Who is

actually going to take a big tech company to court?” she asked, citing cost, time, loss of earnings and potential damage to reputation. “It is just a very impractical way of implementing legislation.”

China is also granting copyrights to AI apps themselves. Why? Because people created those apps. The court held that software alone cannot be the "author" of a work. Chinese copyright law still requires the involvement of a human author to create a copyrightable work.