Permission to Learn:

Modernising Copyright for the AI Age

Onward Corner

FRESHWATER STRATEGY

Contents

Executive Summary 4

Introduction 6

Copyright is for the social good 8

Machine learning does not undermine copyright 9

Input 10

Output 12

The calls and options for reform 14

A restrictive opt-in model would cripple UK AI development 16

An opt-out model has its own dangers 21

Machine-readable opt-out 22

Conclusion – the UK is falling behind 23

Executive Summary

- The UK is a global leader in AI research and development, but outdated copyright rules now threaten its competitiveness and the government's ambition to make Britain an "AI superpower".

- AI training relies on analysing large volumes of publicly available data. This learning process does not reproduce or substitute original works, aligning with the fundamental purpose of copyright: preventing unauthorised copying while enabling creativity and innovation.

- Because UK law lacks a clear text and data mining (TDM) exemption for commercial AI training, major AI developers — including British firms — train their models abroad, reducing domestic investment, talent retention, and AI capability.

- The government has proposed four policy options:

  0) Do nothing

  1) Opt-in licensing

  2) Broad TDM exemption

  3) Narrow TDM exemption (opt-out)

  Only options 2 and carefully designed 3 would meaningfully support UK competitiveness.

- Rights-holder groups advocate an opt-in licensing regime (Option 1), but this would make training advanced AI models in the UK impractical. It would drive AI companies, compute investment, and talent offshore, with little or no benefit to creators.

- Restrictive licensing would not generate meaningful revenue for rights holders, given the high competition among AI models, the low marginal value of any individual work, and the infeasibility of securing permissions at internet scale.

- An opt-in model or a poorly designed opt-out model would risk British users losing access to leading global AI tools if developers choose not to comply with uniquely burdensome UK rules, leaving Britain with poorer quality models and weaker productivity growth.

- The creative industry relies heavily on advanced AI tools for production and distribution. Restricting AI development in the UK would ultimately

harm these sectors by limiting access to the most advanced technologies, for no practical benefit.

- An opt-out TDM exemption (Option 3) could be workable only if designed with minimal compliance burdens, using established standards such as robots.txt. Heavy-handed transparency or database requirements would again drive training offshore.

- International competitors, including the US (under fair use), Japan, Singapore, and the EU (opt-out), already offer clearer, more permissive frameworks for AI training. The UK is therefore falling behind.

- The most effective reform would be a broad, Japan/Singapore-style TDM exemption for commercial purposes, amending Section 29A CDPA. This would give AI developers legal certainty while retaining protections against infringing outputs.

- Without urgent reform, the UK risks becoming a passive consumer of AI innovation developed abroad: its data used elsewhere, its companies and researchers relocating, and its economy missing out on a sector expected to add billions to our economy.

Introduction

The United Kingdom has played a pioneering role in the computing revolution, dating back at least to Alan Turing's groundbreaking work at Bletchley Park. The seismic shift brought about by artificial intelligence (AI) is no exception: from world-leading academic and policy initiatives to pioneering labs like Google DeepMind and Stability AI. This underpinning provides a base for a thriving ecosystem of startups and established firms, while the UK is at the forefront of efforts to harness and integrate AI across the public and private sectors.

The UK is ranked third in Stanford University's Global AI Vibrancy Tool, trailing only the United States and China.[1] This includes a third-place ranking in total AI private investment, which has reached US$28 billion since 2013, and a second-place ranking in the number of newly funded AI firms, with 116 companies in 2024 alone.[2] Britain is in pole position to benefit from higher economic growth: AI could add £550 billion to the UK economy by 2035, alongside breakthroughs in life-saving medical treatments, driverless cars, better public services, and cutting-edge scientific discoveries.[3] The government's AI Opportunities Action Plan, released in January 2025, outlined ambitions to accelerate AI development and adoption, support innovation and attract capital and talent.[4] Prime Minister Keir Starmer has said he wants to make Britain an AI 'superpower'.[5] But nothing is guaranteed.

Governments around the world are investing heavily in AI to secure competitive and strategic advantages. In the global race, policy choices will be decisive. The UK has a solid foundation, but it also has several key challenges. Data centres — essential for AI development and adoption — are being hindered by restrictive planning rules.[6] The UK has the highest industrial electricity costs among developed countries, making energy-intensive industries such AI significantly less competitive.[7]

However, perhaps the gravest present danger and the most significant opportunity for reform lies in the UK's approach to copyright: an outlier among competitor nations. AI models are typically trained on vast amounts of publicly available data to learn patterns, which are represented in mathematical weights, enabling them to generate new content without storing or reproducing the underlying data. However, the UK's copyright regime creates legal uncertainty for AI, prompting developers to train their models in jurisdictions with more permissive laws, such as the US, Japan, and Singapore. Notably, none of the most prominent large language models, like OpenAI's ChatGPT or Anthropic's Claude, are trained in the UK. Even British companies, like London-based Stability AI, have firmly stated that their AI training takes place entirely outside the UK.[8] This situation does not bode well for Britain being an AI superpower.

This is by no means a new or unacknowledged issue. In December 2024, a government consultation proposed a text and data mining (TDM) exception that would explicitly permit commercial AI developers to utilise publicly available content in training models, unless rights holders opt out, to align with the

European Union (EU)'s approach.[9] In response, the rights holders, including news media, book publishers, and the music industry, launched the "Make it Fair" campaign. They advocate for an *opt-in* model, requiring AI developers to pay a licence fee to use materials in AI training, along with cumbersome transparency requirements. Some even advocate for extraterritoriality, meaning that AI models that fail to follow UK-specific rules would be inaccessible in Britain.[10] (This proposal would conflict with Britain's obligations under the Berne Convention, which affirms the territorial nature of copyright, meaning no state may prescribe legal rules for activities occurring beyond its national borders.[11])

These demands risk severe consequences, not only for the UK's AI sector but also for the broader economy and society, with little to no benefit to the creative sector. Changes that make the UK a more difficult place to train and adopt AI than even the EU, let alone the more hospitable jurisdictions, would prompt an exodus of top AI companies, talent, and investment. These proposals are unlikely to result in significant payments to copyright holders. Instead, AI companies will simply continue to train their models elsewhere. This approach risks the UK becoming an AI laggard, with less advanced models available for British consumers across all sectors. Ironically, one of the biggest losers would, in fact, be the creative industry, which is already relying on cutting-edge AI tools to produce world-leading films and music. Areas where the UK is a world leader. Julia Willemyns and Pedro Serôdio of the Centre for British Progress highlight that losing out on the benefits of AI could cost Britain's economy £106-175 billion by 2035, while just a small fraction (£8.3 billion) of the creative sector is even plausibly at any risk from AI open-web learning.[12]

This briefing paper highlights the risks to Britain's AI sector and broader economy from inappropriate copyright policy, underscoring the urgent need for pro-innovation reforms. It begins by revisiting foundational principles about the appropriate role of copyright and the nature of AI. This is followed by a discussion of the reform options and risks associated with an excessively cumbersome regulatory regime.

Copyright is for the social good

Britain has long been a global leader in the development of copyright law. The Statute of Anne, enacted in 1710, was the world's first legislation to grant authors exclusive rights over their works, allowing them to decide how to reproduce, adapt, distribute and display their work. The statute's full title, *An Act for the Encouragement of Learning*, reflects its purpose: to promote the development of knowledge by preventing unauthorised reproduction and ensuring authors are fairly compensated.

Copyright was never intended as an innate, perpetual or even moral right, particularly in the US and UK context.[13] It is a time-limited privilege created by law that seeks to balance incentives for producing creative work with broader access to knowledge and culture.[14] Copyright protects the specific expression of original works, but not a style, as this would prevent learning from and evolving of ideas. There are further exceptions for uses in the public interest, particularly those related to education, learning, and innovation. For example, quoting from one work into another.

The need for copyright to serve a public interest has long been reflected in debates on the topic. Thomas Babington Macaulay (later, Lord Macaulay) classically described copyright in a House of Commons debate as a monopoly.[15] "It is good that authors should be remunerated," he said. "Yet monopoly is an evil... the evil ought not to last a day longer than is necessary." Ben Southwood of the Adam Smith Institute similarly argued that IP rights are valuable not inherently, but because they contribute to human flourishing by incentivising innovation and economic growth.[16] Going one step further, Tom G. Palmer argues that IP rights are morally and practically problematic, as they impose coercive restrictions on the use of texts, which are by nature non-rivalrous and infinitely shareable.[17]

The emergence of any new technology, particularly one as transformative as AI, should prompt policymakers to ask not how to preserve the status quo, but what serves the broader social and economic interest. AI has the potential to deliver breakthroughs across vital sectors, from diagnosing diseases earlier to personalising learning and predicting natural disasters. These advances are already taking shape, from AI-powered tutoring to predictive flood forecasting and early cancer detection. To unlock AI's full potential, the UK's legal framework must reflect the same spirit that animated the Statute of Anne: enabling progress by balancing reward for creators with the broader good of society. Limiting AI's development through overly rigid copyright restrictions would not protect creativity; it would restrict the very conditions under which it flourishes.

Machine learning does not undermine copyright

AI is the field of computer science focused on developing systems that can analyse data, recognise patterns, and generate responses or actions based on that information. Generative AI relies on machine learning, where algorithms are trained on large datasets to identify patterns and make predictions, or create new content based on what it learns. They are not databases, but rather, inference machines built on volumes of data.

Large language models (LLMs), like ChatGPT, are trained by scraping publicly available web content — as well as private datasets, synthetic data and reinforcement learning from human feedback (known as RLHF) — to capture linguistic and conceptual patterns. These patterns are represented as mathematical weights in neural networks, enabling the model to generate contextually relevant content. Importantly, AI models do not store or replicate training data verbatim. Instead, they encode abstract relationships. When generating output (e.g., a poem or image), the model combines these learned patterns to create novel content, rather than copying and pasting from the dataset.

Access to large amounts of data is central to machine learning. Recent advances in AI have been driven by training on enormous datasets, including trillions of texts, images, audio, and video. This enables models to capture the structure and nuance of human language, resulting in emergent behaviours such as reasoning, translation, summarisation, and code generation. The quality of AI models depends on the scale and diversity of their training data.

Input

Copyright law prohibits the *unauthorised reproduction of content*. As AI models learn from but do not reproduce training data, their operation aligns with copyright principles: learning, but not copying. Analogously, it is akin to a person reading a large number of books, synthesising information, and forming and publishing an article on the same topic.

While rights holders argue that AI companies profit from this process, copyright has never barred individuals from benefiting after viewing and being inspired by others' works, such as selling a painting inspired by the style of another artist. Some argue that the sheer scale at which large language models process data disqualifies them from the learning principle. Yet copyright law has never imposed a quantitative threshold beyond which learning ceases to be lawful. There is no reason to believe that a PhD student, who engages with thousands of texts over several years to inform their thinking, is infringing copyright simply by absorbing and synthesising a large amount of material. In fact, the more diverse the input, the less likely that the output will be based on any single source, reducing the risk of copyright infringement. To claim that learning becomes illegal merely by expanding in scale is both conceptually inconsistent and practically unworkable.

Nevertheless, there remains significant legal uncertainty regarding whether this learning principle extends to AI training in the UK. During the machine learning process, for example, a developer may download and temporarily store a local cache of the materials on their system. These copies are not accessible for subsequent reproduction and cannot be referred to by the AI model. Section 28A of the CDPA allows the creation of temporary copies of works during technological processes, provided these copies are transient or incidental, an integral and essential part of the process, have no independent economic significance, and are solely for enabling lawful use of the work.[18] While this provision could cover temporary copies being made during AI training, the question is yet to be tested in the courts and remains contested among experts.[19] This is further complicated by Section 29A of the CDPA, which explicitly permits text and data mining (TDM) for *non-commercial* research where the user has lawful access.

Taken to its logical extreme, a reductive interpretation of reproduction would capture almost any act of data processing. On this view, even the fleeting imprint of information held for a nanosecond in a microchip's transistors could be construed as storage. This kind of reasoning has surfaced in other contexts, beyond AI, whenever the law grapples with what constitutes a copy in the digital age. There is, therefore, considerable merit in rejecting such overreach and instead grounding the analysis in the original purpose of copyright rules: to prevent unauthorised reproduction and exploitation that undermines creators, not to penalise transient, incidental processes undertaken without malicious intent.

In practice, however, the lack of an explicit TDM exemption for commercial purposes makes training advanced AI models both legally uncertain and commercially risky, effectively deterring developers from conducting such activities in the UK. As a result, industry-leading models, including those developed by British companies, are being trained in jurisdictions with clearer and more permissive frameworks, such as the United States, Japan, and Singapore. This leaves the UK at a significant competitive disadvantage in the global AI race, meaning models are not optimised for UK users or reflect British values.

Output

AI models are designed to produce novel outputs, drawing on patterns learned during training and incorporating elements of randomness or probability in their generation process, meaning that the same prompt can yield different results each time the AI model is prompted. In human terms, AI is 'standing on the shoulders of giants' — output is influenced by what has come before while creating something new. This process does not undermine the market of the underlying training data. That is, AI outputs are not a substitute for the underlying works: those who want to read a novel, watch a film, listen to a song, or read a news article must still obtain the original.

AI developers have also introduced layers of safeguards to prevent their models from simply regurgitating specific works. This includes output filtering and posttraining evaluation. AI companies have every incentive to avoid clear violations of copyright and the associated legal risks.

An AI model may, in some cases, generate outputs that resemble over-represented or frequently repeated items in its training data, particularly when intentionally prompted. Such similarities arise from the model's statistical learning of patterns in that data. This gives rise to novel legal questions. In the first instance, if such resemblance occurs in a limited or incidental way – akin to a book author quoting and attributing a few lines from a newspaper article – it is highly unlikely to constitute a violation of UK copyright law. This is because the partial use of publicly available copyrighted material for purposes such as criticism, review, and quotation is permitted under the fair dealing provision (Section 30 of the CDPA). There is also explicit protection for parody, caricature, and pastiche (Section 30A of the CDPA), which includes work created in the style of another's work. This is likely to encompass images in the style of, for example, Studio Ghibli.[20]

On the other hand, questions may arise under copyright law when an AI model generates output that is identical or similar to a substantial part of a protected work, such as several paragraphs of a news article, without proper attribution. Whether such output constitutes reproduction, adaptation, or a new work remains unsettled in law. There has been some discussion about whether the law should be updated to reflect 'best efforts' to avoid infringing outputs. In practice, assessing whether there has been a violation requires considering the nature and extent of the similarity, the circumstances in which it arose, and how closely it aligns with established copyright tests. In any case, it is in the output of models that copyright implications can best be reasonably assessed, rather than in the input to models.

Copyright prevents the unauthorised reproduction of works to encourage innovation and creativity. It is not intended to protect a style or idea, nor to undermine creativity, innovation or the broader public interest. AI training is a form of learning that may involve temporary, incidental copying, just as people do

when learning from a book. Arguably, this is already allowed under UK law. But, in practice, legal uncertainty is preventing advanced AI models from being trained in the UK.

It is essential to distinguish between inputs, the learning process, which does not breach copyright principles, and outputs, where assessment is relevant, even if the legal position on such cases remains unclear.

Filmmaker James Cameron likens human learning to AI models, noting that, like screenwriters drawing inspiration from past works, we absorb data and must transform it into something sufficiently original to create independently:

> "So I think the whole thing needs to be managed from a legal perspective at what's the output, not what's the input. You can't control my input. You can't tell me what to view, and what to see and where to go. My input is whatever I choose it to be and whatever has accumulated throughout my life. My output, the script I write, should be judged on whether it's too close, too plagiaristic."[21]

The following sections examine the government's proposed reform options and their associated implications.

The calls and options for reform

In 2011, the government-commissioned Hargreaves Review of Intellectual Property concluded that outdated copyright rules had failed to keep pace with technological and social change. Hargreaves warned that "the law can block valuable new technologies, like text and data mining, simply because those technologies were not imagined when the law was formed."[22] The review recommended a text and data mining exemption to enable machine learning, but this was ultimately only introduced for non-commercial purposes.

In practice, the non-commercial exemption has proven to be of limited utility, as even publicly funded and non-profit institutions, such as the NHS, routinely collaborate with private sector partners, and universities increasingly seek to commercialise their research. If the intention was to drive public benefit from AI and machine learning, the restriction has undermined it, because real-world innovation, scale, and delivery to consumers typically require private sector investment. In a consumer-driven economy, public utility is maximised not by limiting commercial application, but by encouraging it.

There have been several further attempts at reform. In 2022, a government consultation led by the Intellectual Property Office proposed a copyright exception for TDM for any purpose. The extension of the TDM was also backed by policy reviews undertaken for the current and previous governments, including Sir Patrick Vallance's Pro-innovation Regulation of Technologies Review in 2023 and Matt Clifford's AI Opportunities Action Plan in January 2025. "The current uncertainty around intellectual property (IP) is hindering innovation and undermining our broader ambitions for AI, as well as the growth of our creative industries," Clifford warned. "The UK is falling behind."[23]

In response to these calls, the government's Copyright and Artificial Intelligence consultation, published in December 2024, proposed the following options:

- 0: Doing nothing;
- 1: Copyright licensing requirements (opt-in model);
- 2: A broad text and data mining (TDM) exemption (no opt-out); or
- 3: A narrow text and data mining (TDM) exemption (opt-out model).[24]

Option 0: Doing nothing

Inaction would mean it would remain too legally risky to train leading AI models in the UK, resulting in top talent and investment being pulled away. It would continue to be possible to access AI tools. But even this would be less than guaranteed. The ongoing *Getty Images vs. Stability AI* case illustrates the legal uncertainty. If the High Court decides to apply UK copyright law to AI models trained abroad, British users may lose access to leading tools such as ChatGPT and Google Gemini.[25] In this scenario, UK copyrighted material would still be used to train AI; it would just happen elsewhere for the benefit of others. The loss of access to leading AI tools would have broader economic significance for the UK's productivity – the equivalent of sticking with horse and cart while others adopt

the combustion engine. In summary, inaction would fail to make the UK an 'AI superpower' while imposing legal risks down the track, depending on how courts opt to interpret the existing law.

Option 1: Copyright licensing requirements (opt-in model)

Option 1 is an "opt-in" model that would require AI developers to obtain explicit permission or licences from copyright owners before using material for training purposes. This option, preferred by rights holders, would make it logistically and financially unviable to build large-scale AI models domestically, driving away talent, investment, and innovation. This would undermine the UK's ambition to lead in AI and could have serious downstream consequences for start-ups, businesses adopting AI, and even the creative sector itself, which relies heavily on access to cutting-edge AI tools. The dangers of this approach are discussed further in the next section.

Option 2: A broad text and data mining (TDM) exemption (no opt-out)

Option 2 is a broad text and data mining (TDM) exemption that allows AI developers to use publicly available copyrighted works for training without permission. This would provide the legal clarity and freedom necessary to enable the development and use of AI in the UK. It would ensure that developers can access the vast, diverse datasets required to train effective and competitive models without facing prohibitive licensing costs or logistical barriers. This option would align UK with more innovation-friendly jurisdictions, such as Japan and Singapore, and the US under fair use, helping to attract talent, investment, and infrastructure while avoiding the legal uncertainty and complexity associated with opt-in or opt-out systems. Crucially, it would still preserve protections against output-based copyright infringement and does not interfere with deals for private or high-value datasets. This option maximises innovation and economic benefit without compromising core copyright principles.

Option 3: A narrow text and data mining (TDM) exemption (opt-out model)

Option 3 is a narrow text and data mining (TDM) exemption that allows AI developers to use publicly available copyrighted works for training only if rights holders have not explicitly opted out using a recognised mechanism. This is intended to replicate the EU's approach, although, as will be discussed, this very much depends on technical issues such as machine-readable opt-outs and transparency requirements. The government stated in its consultation that this was its preferred option; however, subsequent reporting indicates otherwise.[26] This may be a workable second-best option for the UK, if implemented proportionally. However, it would fail to provide the UK with a competitive advantage internationally and is therefore unlikely to attract significant inward investment, as it would still be legally more straightforward and financially less costly to train AI in other jurisdictions.

A restrictive opt-in model would cripple UK AI development (Option 1)

Britain already has a more restrictive copyright regime than its major competitors. The EU allows AI training with an 'opt-out' by rights holders, Japan and Singapore have no limits on AI training using publicly available information, while the United States, under 'fair use' principles, provides for training on any publicly available data. As a result, AI models are typically trained outside of the UK.

But the situation could worsen under proposals being put forward by some rights holders. The 'Make it Fair' campaign claims that AI models are 'stealing' copyrighted work, devaluing their labour and eroding traditional business models. In response, they are calling for the introduction of an 'opt-in' system, whereby AI companies are required to obtain explicit permission, including licensing agreements involving payments, before training. AI models not trained under these rules could also be made inaccessible to British users.

The campaign's argument may have some instinctive appeal, and it is easy to understand the motivations of some rights-holders who believe AI undermines their business model. At its core, however, this appears to be an attempt to carve out a new revenue stream under the guise of fairness, an attempt at 'rent-seeking'. This is hardly surprising; there's every incentive to test how far they can push for a revenue transfer, especially when there's little downside to trying.

The arguments for this model, however, fail at both a principled and practical level. As discussed, AI is not 'copying' material into its outputs; it is, like humans do, learning from existing materials and producing new outputs. Beyond this, a more restrictive regime proposed by rights holders could have severe consequences for Britain's AI future, with limited or no benefits for the creative sector.

AI companies, talent and investment will go offshore, the UK economy will suffer

If the UK becomes a more difficult environment than other countries for training and deploying AI, leading companies, top talent, and investment are likely to leave. Those who stay, notably smaller homegrown developers, will face a growing disadvantage, as AI continues to be trained overseas on UK content without licensing. Rather than becoming an 'AI superpower,' Britain would become an AI backwater, with serious economic repercussions. The impact would extend beyond foundational model developers to include start-ups building AI applications and millions of UK businesses that stand to gain from integrating AI into their operations. A less attractive AI ecosystem would also discourage investment in domestic computing infrastructure, further weakening downstream sectors, including the creative industries.

The government's consultation warns that the rights holders' preferred licensing approach is "highly likely to make the UK significantly less competitive compared

to other jurisdictions, such as the EU and US, which do not have such restrictive laws".[27] A survey of 500 developers, investors and others working in AI in the UK, undertaken by J L Partners for the Computer & Communications Industry Association, found that 99% of AI developers say AI development is reliant on text and data mining using publicly available information and three-quarters said a failure to introduce a text and data mining (TDM) exemption would signal that the UK is less competitive. The UK is already the biggest exporter of AI talent to the USA.[28] Adding further regulatory obstacles would only compound the loss of talent and economic opportunity.

If the policy were to include blocking access to AI models not trained under UK copyright rules from being used by British users, it could amount to an effective import ban on leading models. This is because AI developers will likely be willing to forgo deployment in the UK altogether to avoid restrictive regulations and legal risks. After all, the UK accounts for less than 4% of the global economy and less than 1% of the global population. The likes of OpenAI and Meta have already delayed AI product releases in the UK and the EU over regulatory concerns.

Therefore, Britain not only risks missing out on AI training but also on deployment, further entrenching its disadvantage in the broader UK economy, reducing access to productivity and growth-enabling new technologies. The government's consultation warns that it may be difficult to restrict models trained by other standards from the UK, particularly since cutting-edge models can now be run on a personal laptop. But to the extent that it is successful, there would "be a risk that some of the most capable AI models would not be available in the UK," according to the consultation, "This would significantly limit innovation, consumer choice, and wider benefits of AI adoption across the UK economy."[29] So, AI models would still likely be trained on UK content, from overseas, but nobody in Britain would have access to the benefits. Or, if the law went further and sought to regulate AI outputs generated by models that did not comply with UK copyright rules, it could result in banning a Hollywood film that used AI in production or prohibiting a novelist who relied on AI as a writing companion. This impractical and absurd outcome would ultimately disadvantage both British consumers and creators alike.

The quality of AI itself could also be undermined, risking unintended social harms. Amanda Levendowski has argued that a permissive approach to data access is "quite literally, [necessary for] promoting fairer AI." Without diverse and representative data, models are more likely to amplify harmful stereotypes and exclusionary assumptions. When developers are forced to rely on narrow, legally safe datasets, such as public domain or Creative Commons materials, training data can become skewed toward white, Western, and male-centric perspectives. This compromises both the accuracy and the inclusivity of AI systems.

Limiting access to UK data also means that these biases may be reinforced elsewhere, in models trained abroad that do not accurately reflect British society, values, or cultural nuances. If the UK government is serious about its soft power ambitions and wants its values to be represented in AI tools, it needs to support

the development of the industry in the UK, rather than restricting it to other economies. Britain possesses immense cultural capital, embodied by institutions such as the BBC, the British Council, and the British Library. Without British perspectives in global AI tools, the UK's cultural voice may become diminished, regardless of its soft power initiatives.

The creative sector would not benefit

Rights-holders are imagining that if the legal regime is changed to require explicit licensing, beneficial monetary agreements will be reached. But this is highly unrealistic. Foundational large language models are a highly competitive field. The entry-level versions of ChatGPT and Claude are already free, along with the emergence of various open-source models like Meta's Llama and DeepSeek. In the competitive AI market, there is little to no revenue to be shared with the creative sector for accessing publicly available information.

Deals have and will continue to be reached to access specific private datasets. For example, the University of Oxford and OpenAI have agreed to digitise 3,500 dissertations from the 15th to 19th centuries, which will presumably be included in future training.[30] There will also continue to be deals for specialist AI applications that require access to private data for fine-tuning and retrieving relevant, up-to date information to generate accurate, domain-specific outputs (e.g. a legal tool drawing on a proprietary legal database). However, the imposition of a licensing regime will not motivate AI developers, who face heavy competition from peers, to begin training their large language models in the UK and pay for the ability to do so. If there is no uptick in AI training in the UK, then there are unlikely to be envisaged deals with the creative sector.

In any case, enforcement would be virtually impossible. Bad actors would continue training on UK-origin content without detection, while compliant firms would bear the burden of costly restrictions. As it stands, it is exceedingly difficult to determine whether opt-out content has been included in a training dataset (as will be discussed further below in transparency requirements). Without robust detection methods, legal rules risk punishing law-abiding companies while doing little to prevent actual misuse.

Even if large AI developers could shoulder the financial burden and enforcement were possible, the practicalities of striking deals would in themselves be a significant barrier. Large language models require vast datasets, and obtaining permissions from billions of individual sources would be technically infeasible and logistically unmanageable.[31] In an opt-in model, most works would be excluded from training, not because rights holders made an active choice, but because the content is old, ownership is unclear, or it was never commercially managed in the first place. An old blog, for example. Meanwhile, transaction costs in copyright licensing markets risk being so high that they outweigh the low value of training data per work, rendering licensing markets extremely difficult, if not impossible, to function. To the extent that a market emerged, there would also be ongoing disputes and associated significant compliance burdens. Therefore, rights

holders' preferred option would make building AI models in the UK not only significantly more expensive but effectively impossible.

The creative sector needs diverse AI to succeed

The loss of AI expertise in Britain would have downstream implications for the creative sector. Rights holders have highlighted statistics showing the creative industries added £124 billion in Gross Value Added (GVA) in 2022.[32] However, over 40% of that figure is comprised of IT, Software, and Computer Services, which would be at the highest risk of losing AI talent and investment. For the other parts of the sector, like film and television, they would find that UK content would still be used in training offshore, but they would receive fewer benefits from more advanced AI tools. Without the latest AI tools, British creatives would find themselves at a competitive disadvantage.

Some rights holders feel threatened by AI. However, the long-term success of the creative industry depends on its ability to adapt and evolve, rather than trying to block the development of AI. From the invention of photography, which sparked fears about the future of portraiture, to the rise of streaming platforms, creative industries have faced technological shifts for centuries. Historical experience demonstrates that new technologies augment existing producers and boost productivity, creating opportunities for higher-quality work and better-paying jobs than existed previously. Those who embraced change found new ways to thrive because innovation and creativity are mutually reinforcing.

There is no credible evidence that AI is undermining human creativity; rather, the opposite is true. AI may be very effective at some tasks, but there is still an underlying demand for the authentic work of authors, musicians, and directors. AI is therefore most likely to play an augmenting rather than a replacing role in most cases. In countries with permissive AI training laws, such as the United States, Japan and Singapore, creative sector output continues to grow.[33] Globally, the same is the case, with record numbers of movies, music tracks, and games being released in recent years, along with the acceleration of AI.[34]

AI is becoming an increasingly important part of the creative process. Media companies have been relying on AI to automate content recommendation and target advertising for several years, and they are now using generative AI in film production, like *The Brutalist*.[35] Smaller and independent creative houses are adopting AI to compete with higher-budget productions.[36] AI can also be used for many functions that expand the potential market for innovative products, such as providing higher-quality and lower-cost voice translations. The arts and media sector is the second-largest user of Claude.[37] A Google survey found that 72% of media and entertainment companies using generative AI reported a positive return on investment.[38] Artist Lex Fefegha, for example, has been exploring interactive play and storytelling powered by AI.[39] Fefegha's Thames Path 2040 uses AI tools to imagine what will remain if rainfall leads to flooding on the Thames Path in 2040.

On the other hand, if the industry somehow becomes dependent on transfer payments from AI companies, as they are currently seeking, there will be reduced pressure to innovate, utilise cutting-edge tools, and develop new business models. The risk is of repeating the experience of the news publishing industry, whose attempt to rely on external subsidies dulled incentives for transformation and left outlets struggling to adapt to the digital age.[40] Historically, the closest equivalent is to the Luddites, who may not have stopped the Industrial Revolution by destroying factory equipment, but did slow down the ability to adapt and grow in the new world.

Similarly, limiting AI or requiring transfer payments would slow down the adoption of innovative new methods and technologies, making it more difficult for the UK to remain competitive. The future of the UK's creative sector does not depend on restricting the ability of AI to train on publicly available content, at little to no benefit to the creative sector or AI developers, but rather on their ability to adapt in the face of ongoing technological and societal pressures.

An *opt-out* model has its own dangers (Option 2)

The government's original preferred option was to create a text and data mining exemption while allowing rights holders the ability to 'opt out' from training. This is intended to align the UK with the EU's approach, though, depending on the details, it could be more burdensome. Julia Willemyns of the Centre for British Progress warns that even the opt-out model would still risk making it relatively more expensive to train in the UK compared to other jurisdictions.[41] According to Willemyns, this would lead to AI investment being directed elsewhere, resulting in a loss of at least £29.9 billion to the UK's GDP over the next five years. If the government proceeds with the opt-out model, it must be carefully designed to minimise compliance burdens and avoid the negative consequences outlined above.

Machine-readable opt-out

An opt-out system requires clear mechanisms that allow rights holders to exclude their content from AI training. The current industry standard, which major AI developers recognise, is robots.txt. This is a widely used machine-readable mechanism that instructs web crawlers not to visit a site, dating back to the early days of the internet. It is a logical starting point for any mandate. The likes of the *BBC* and *The New York Times* are already using the protocol to deny access to their websites for AI training. There are other standards, such as TDMRep, the International Standard Content Code (ISCC), and C2PA, which could also provide machine-readable opt-outs. However, further development and industry-agreed standardisation will be necessary for their adoption.

Alternative approaches, such as mandating an opt-out technology that does not yet exist or has not yet been proven at scale, like content-level opt-outs, would risk creating a unique regulatory burden, thereby putting the UK at a competitive disadvantage akin to adopting the opt-in model. There have been efforts to develop a centralised database containing a list of items that have been opted out of training. It may also be possible in future to use AI itself to initiate requests for content removals. However, this technology is still in its infancy, and a burdensome regulatory approach could result in AI being trained outside the UK once again. A requirement to allow rights holders to withdraw permission retroactively would also be technically infeasible, as an AI model to 'unlearn' content after training.

Minimally burdensome transparency requirements

Rights holders' advocates have sought to require AI developers to disclose the materials they use to train their models, to identify potential infringement. There is already disclosure of the general sources of information used in AI training, including the types of data, by major AI companies. Baroness Kidron's data bill amendment, however, went further by demanding the full publication of all URLs and data used at all stages of training.[42] Such a list would be costly and complex to develop and maintain, and would undermine the proprietary nature of AI model training, making it practically infeasible. It would also be based on a false premise since most of the data scraped for AI training is not used in the training, nor does it contribute to the model's output. Publishing a list won't allow you to determine what has actually influenced the model, because it's the weighting and interplay of vast amounts of data (the proprietary process of the model), not individual items, that shapes the output.

If the government decides to proceed with transparency requirements, it will be essential that these are proportionate, that they limit the compliance burden and do not undermine trade secrets. This is another area in which the industry is likely to develop further standards over time, rather than one that requires immediate and prescriptive legislative intervention. It is also likely that if the UK imposes particularly burdensome requirements, AI developers will remove access to their models for UK users. There have already been several cases of AI product releases being delayed for European users due to regulatory risks, including Meta AI, OpenAI's Sora, and, most recently, Apple Intelligence.

Conclusion – the UK is falling behind

The UK is at a crucial point in the global race to lead in AI. While it boasts a worldclass AI ecosystem and government ambitions to become an AI "superpower," outdated copyright laws threaten to hinder its potential. Several of Britain's international peers have already taken steps to offer legal clarity and certainty to AI developers. If the UK does not do the same, it risks falling behind in a field that will define economic, technological, and geopolitical strength over the coming decades.

The United States, Japan, Singapore, and the European Union provide supportive legal frameworks for AI training. The US depends on the fair use doctrine, which courts have interpreted favourably in landmark cases involving large-scale text and data analysis. A court ruling in the United States in June 2025 determined that Anthropic can train its AI using published books without the authors' consent, provided it falls under fair use principles. "Authors cannot rightly exclude anyone from using their works for training or learning as such," the judge concluded: "Everyone reads texts, too, then writes new texts. They may need to pay for getting their hands on a text in the first instance. But to make anyone pay specifically for the use of a book each time they read it, each time they recall it from memory, each time they later draw upon it when writing new things in new ways would be unthinkable. For centuries, we have read and re-read books. We have admired, memorized, and internalized their sweeping themes, their substantive points, and their stylistic solutions to recurring writing problems."[43]

The EU, although more regulated, allows commercial AI training by default through its opt-out system. Japan has gone further, with statutory reforms that explicitly permit the use of copyrighted materials for AI training, as long as the use is analytical and non-expressive. Singapore's 2021 reforms repeat this openness by offering a clear, unambiguous exemption for computational data analysis.

These countries understand that AI is a general purpose technology, akin to electricity or the internet. Their legal systems recognise that training AI models on publicly accessible data does not undermine the creative sector. It empowers it, fosters innovation, and ensures competitiveness in the digital age.

By contrast, the UK's copyright framework remains ambiguous in its treatment of commercial AI training. The current law fails to clearly distinguish between learning (input) and reproduction (output), and provides no explicit exemption for commercial text and data mining. This uncertainty deters developers from training models in the UK, pushing innovation, infrastructure, and investment offshore.

The government's proposed opt-out model may be a step forward, but it lags behind the more permissive regimes in Asia and North America. Without careful

design to minimise compliance burdens, even this could stifle development. Worse still, the adoption of a restrictive opt-in licensing regime, favoured by some rights holders, would render it practically and financially unfeasible to build large-scale AI models domestically.

The best course for the UK is to adopt a Japan or Singapore-style approach, amending Section 29A of the CDPA to extend the existing text and data mining exception to commercial purposes. This would provide developers with the certainty to innovate while preserving copyright protections for outputs. Such a regime aligns with existing legal and moral principles and would help attract global AI talent and investment to the UK.

If the UK is serious about its goal of becoming an AI superpower, it must act now. Trying to strike a 'balanced' position that imposes unique compliance burdens is more likely to backfire, leaving the UK isolated and uncompetitive. Instead, the government must embrace a future-facing copyright regime that empowers developers, protects creators through output regulation, and ensures the UK remains a hospitable environment for AI advancement.

There are several historical parallels of UK governments making difficult regulatory decisions that ultimately enhanced Britain's global competitiveness: the ‘Big Bang’ financial reforms transformed London into a world-leading financial centre; telecoms liberalisation opened the door to mobile and internet innovation; and independent broadcasting rules fuelled the launch of new radio and television channels, enabling a globally successful creative sector.

The choice now is stark: lead or lag. Without reform, the UK will become a spectator in the AI revolution: its content being used to train models abroad, its innovators forced to build elsewhere, and its economy left trailing behind more ambitious countries.

Endnotes

[1] Stanford Institute for Human-Centered Artificial Intelligence, *Global Vibrancy Tool*, https://hai.stanford.edu/ai-index/global-vibrancy-tool (2025).

[2] Stanford Institute for Human-Centered Artificial Intelligence, *2025 AI Index Report*, https://hai.stanford.edu/ai-index/2025-ai-index-report, p.252 and p.255 (2025).

[3] Microsoft UK, *AI Could Boost UK GDP by £550 Billion by 2035, Research Shows*, https://ukstories.microsoft.com/features/ai-could-boost-uk-gdp-by-550-billion-by-2035research-shows/ (2025).

[4] UK Government, *AI Opportunities Action Plan*, https://www.gov.uk/government/publications/ai-opportunities-action-plan/aiopportunities-action-plan (2025). Reuters, *UK PM Starmer to Outline Plan to Make Britain World Leader in AI*, https://www.reuters.com/world/uk/uk-pm-starmer-outline-plan-make-britain-worldleader-ai-2025-01-12/ (2025).

[5] Data Centre Review, *Data Centre Planning Reform*, https://datacentrereview.com/2025/04/data-centre-planning-reform/ (2025).

[6] UK Government, *International Industrial Energy Prices*, https://www.gov.uk/government/statistical-data-sets/international-industrial-energyprices (2025).

[7] Pinsent Masons, *Getty Images v Stability AI: Implications for Copyright Law and Licensing*, https://www.pinsentmasons.com/out-law/analysis/getty-images-v-stability-aiimplications-copyright-law-licensing (2024).

[8] UK Government, *Copyright and Artificial Intelligence Consultation*, https://www.gov.uk/government/consultations/copyright-and-artificialintelligence/copyright-and-artificial-intelligence (2024).

[9] UK Parliament, *Baroness Kidron's Amendment 49D, Data (Use and Access) Bill (HL)*, https://hansard.parliament.uk/lords/2025-05-19/debates/AD1D6032-D2DF-453B-B4A787FFE6ED1428/Data(UseAndAccess)Bill(HL) (2025).

[10] In the leading treatise on the Berne Convention, Professors Ricketson and Ginsburg open the foreword with: "Copyrights are territorial: they are rights conferred by the laws of nation states which have effect within their national territories." (Sam Ricketson and Jane C. Ginsburg, International Copyright and Neighbouring Rights: The Berne Convention and Beyond, 3rd edn (Oxford UP, 2022), at v.) According to Paul Goldstein and P. Bernt Hugenholtz (International Copyright: Principles, Law, and Practice, 4th edn 87 (2019)), international copyright law defines the "territoriality principle" as holding "that a state has no competence to prescribe legal rules to govern activities that occur outside its national borders." The treatise explains the two bases for the principle, namely, first, national sovereignty and, second, predictability of international commerce "by securing reasonable investment expectations."

[11] Willemyns, Julia and Serôdio, Pedro, *Who Actually Benefits from an AI Licensing Regime?*, https://britishprogress.org/reports/who-actually-benefits-from-an-ai-licensing-regime (2025).

[12] See A Balance of Interests: the Concordance of Copyright Law and Moral Rights in the Worldwide Economy Michael B. Gunlicks Spotts Fain Chappell & Anderson, P.C

[13] Omaar, Hodan, *Selective Outrage Over AI and Copyright*, https://britishprogress.org/ukday-one/copyright-ai-the-case-for-a-pro-growth-approach (2025).

[14] The Public Domain, *Macaulay on Copyright*, https://www.thepublicdomain.org/2014/07/24/macaulay-on-copyright/ (2014). [16] Southwood, Ben, *Patently Good: A Defence of Intellectual Property*, Adam Smith Institute, https://www.adamsmith.org/research/patently-good-a-defence-of-intellectual-property

[17] Palmer, Tom, *Are Patents and Copyrights Morally Justified? The Philosophy of Property Rights and Ideal Objects*, Harvard Journal of Law & Public Policy, Volume 13, Number 3, http://tomgpalmer.com/wp-content/uploads/papers/palmer-morallyjustified-harvardv13n3.pdf (1990).

[18] Lexology, *AI and Copyright: Legal Considerations*,

https://www.lexology.com/library/detail.aspx?g=74ab3ff0-d02f-4e01-93fa-d6935d53e3ab [19] Andres Guadamuz, the Reader in Intellectual Property Law at the University of Sussex and the Editor in Chief of the Journal of World Intellectual Property, has argued that "most inputs may fall under existing exceptions and limitations," see https://academic.oup.com/grurint/article/73/2/111/7529098. By contrast, Dr Hayleigh Bosher, Reader in Intellectual Property Law and Associate Dean at Brunel University London, argues that the 28A exception is too narrow to apply, as the storage is not transient or incident and nor is it of no economic significance, see https://committees.parliament.uk/writtenevidence/126844/html/.

[20] JJ Taylor Solicitors, *AI-Generated Art: Navigating Copyright and Ethical Boundaries*, https://jjtaylorsolicitors.com/ai-generated-art-navigating-copyright-and-ethicalboundaries/

[21] YouTube, *AI and Copyright Discussion*, https://www.youtube.com/watch?feature=shared&t=1903&v=qOdjM14QW0s

[22] UK Government, *IP Review Final Report*, https://assets.publishing.service.gov.uk/media/5a796832ed915d07d35b53cd/ipreviewfinalreport.pdf

[23] UK Government, *AI Opportunities Action Plan*, https://assets.publishing.service.gov.uk/media/67851771f0528401055d2329/ai_opportunities_action_plan.pdf (2025).

[24] UK Government, *Copyright and Artificial Intelligence Consultation*, https://www.gov.uk/government/consultations/copyright-and-artificialintelligence/copyright-and-artificial-intelligence (2024).

[25] There are also several other issues in the case, such as trademark and passing off claims, for further discussion, see Pinsent Masons, *Getty Images v Stability AI: Additional Claims*, https://www.pinsentmasons.com/out-law/analysis/getty-images-v-stability-aiimplications-copyright-law-licensing (2024).

[26] The Guardian, *Ministers Face UK Copyright and Artificial Intelligence Vote in Parliament*, https://www.theguardian.com/technology/2025/may/04/ministers-uk-copyrightartificial-intelligence-parliament-vote (2025).

[27] UK Government, *Copyright and Artificial Intelligence Consultation*, https://www.gov.uk/government/consultations/copyright-and-artificialintelligence/copyright-and-artificial-intelligence (2024).

[28] Zeki Data, *The State of AI Talent* 2024, https://zekidata.com/report-excerpt/the-state-ofai-talent-2024/ (2024).

[29] UK Government, *Copyright and Artificial Intelligence Consultation*, https://www.gov.uk/government/consultations/copyright-and-artificialintelligence/copyright-and-artificial-intelligence (2024).

[30] University of Oxford, *Oxford and OpenAI Launch Collaboration to Advance Research and Education*, https://www.ox.ac.uk/news/2025-03-04-oxford-and-openai-launchcollaboration-advance-research-and-education (2025).

[31] There are some emerging efforts to address this technical problem, like CloudFlare's 'pay per crawl,' but mandating this sort-of approach for all web crawling would still raise issues for sites that fail to opt-in and cases where ownership is unclear.

[32] UK Government, *DCMS Sectors Economic Estimates: Regional Gross Value Added 2022*, https://www.gov.uk/government/statistics/dcms-sectors-economic-estimates-regionalgva-2022/dcms-sectors-economic-estimates-regional-gross-value-added-2022 (2024). [33] RIAA, Grand View Research, CCIA, Deloitte, *Multiple Sources on Creative Economy*, https://www.riaa.com/2024-year-end-music-industry-revenue-report-riaa/, https://www.grandviewresearch.com/horizon/outlook/movie-and-entertainmentmarket/united-states, https://ccianet.org/wp-content/uploads/2024/01/CCIA_Copia_Sky-Is-Rising-2024-Edition_1pager.pdf, https://www.deloitte.com/uk/en/Industries/tmt/perspectives/the-future-of-thecreative-economy.html

[34] CCIA, *The Sky Is Rising: 2024 Edition*, https://ccianet.org/wpcontent/uploads/2024/01/CCIA_Copia_Sky-Is-Rising-2024-Edition_1pager.pdf (2024).

[35] Bertelsmann, Arthur D. Little, and Enders Analysis, *State of Play: Exploring Generative AI's Transformative Effects on the Media & Entertainment Industry*, https://www.bertelsmann.com/media/news-und-media/downloads/bertelsmannstateofplay-genai.pdf

[36] Aapti Institute and Microsoft, *AI and Economic Impact*, https://aapti.in/wpcontent/uploads/2024/07/Aapti-x-Microsoft-28-Jun-6-pm-compressed.pdf

[37] Arts and media account for 10.3% of all Claude conversations, despite comprising just 1.4% of the overall US economy, see https://www.anthropic.com/news/the-anthropiceconomic-index

[38] Google Cloud, *Media and Entertainment: Gen AI ROI Report*, https://cloud.google.com/transform/media-entertainment-gen-ai-roi-report-dozenreasons-ai-value

[39] Lex Fefegha, *Personal Website*, https://lexfefegha.com/

[40] Institute of Economic Affairs, *Breaking the News: Should Digital Platforms Be Required to Fund News Publishers?*, https://iea.org.uk/publications/breaking-the-news-should-digitalplatforms-be-required-to-fund-news-publishers/

[41] British Progress, *Copyright & AI: The Case for a Pro-Growth Approach*, https://britishprogress.org/uk-day-one/copyright-ai-the-case-for-a-pro-growthapproach (2025).

[42] Newton-Rex, Ed, *The Insurmountable Problems with Generative AI Opt-Outs*, https://static1.squarespace.com/static/5cc5785816b6406e50258c5c/t/67368c12cc35b5469 feb0bfd/1731628050768/The+insurmountable+problems+with+generative+AI+opt-outs.pdf (2024).

[43] CourtListener, *Bartz v. Anthropic PBC*, https://www.courtlistener.com/docket/69058235/bartz-v-anthropic-pbc/

Onward Corner

FRESHWATER STRATEGY