



ForesightLAB



Copyright and Artificial Intelligence

A Foresight Lab literature review

March 2025

Dr Ekaterina Tarnovskaya and Prof Graham Hitchen



THE UNIVERSITY
of EDINBURGH



Loughborough
University

Goldsmiths
UNIVERSITY OF LONDON

The CoSTAR Foresight Lab

Driven by the UK's leading Creative Industries experts, the [CoSTAR Foresight Lab](#) is researching the adoption, use and impact of new, emergent and convergent technologies in gaming, TV, film, performance and digital entertainment.

Our findings will inform research, development and innovation across the Creative Industries, including the R&D taking place through the convergent screen technologies and performance in real time (CoSTAR) programme, the UK R&D network for creative technology.

[CoSTAR](#) is a £75.6 million national R&D network of laboratories that are developing new technology to maintain the UK's world-leading position in gaming, TV, film, performance, and digital entertainment. Delivered by the UKRI Arts and Humanities Research Council, the programme is supporting new innovations and experiences that will enrich the UK's creative industries, economy, and culture. The network comprises the National Lab, the Realtime Lab, the Live Lab, the Screen Lab and the Foresight Lab. CoSTAR is funded through UK Research and Innovation's Infrastructure Fund, which supports the facilities, equipment and resources that are essential for researchers, businesses, and innovators to do groundbreaking work. You can find out more by visiting www.costarnetwork.co.uk.

Foresight Lab partners and contributors

Core Partners:

Prof Jonny Freeman; Jac Sanscartier (Goldsmiths, University of London)
Prof Frauke Zeller; Prof Melissa Terras; Dr Suzanne R Black (University of Edinburgh)
Prof Graham Hitchen; Dr Ekaterina Tarnovskaya; Dr Vicki Williams (Loughborough University)
Rishi Coupland; John Sandow; Brian Tarran (BFI)

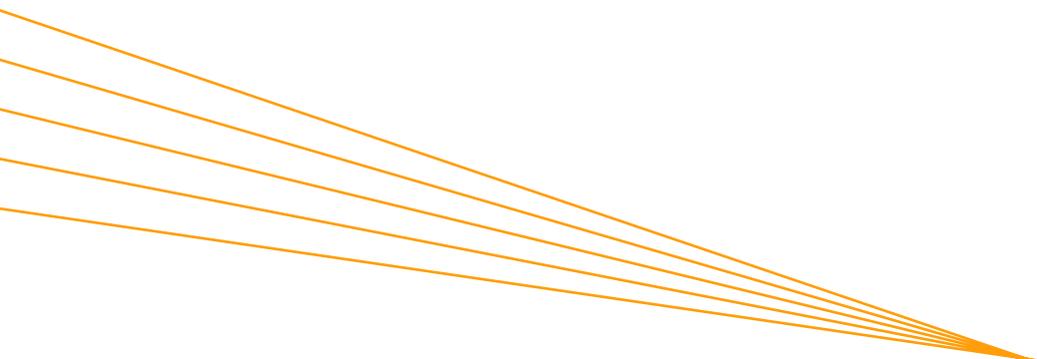
Delivery Partners:

Florence Mansfield (Arup)
Prof Hasan Bakhshi MBE; Dr Tom Cahill-Jones; Prof Giorgio Fazio; Dr Sarah Najm (Creative Policy and Evidence Centre)
Brendan Miles (Data Thistle)
Chiara Badiali; Natalie Highwood; Alison Tickell (Julie's Bicycle)
Noemi Ponzoni; Anna Stewart; Lewis Turner-Brown (i2 media research)
Leon Forde; Marta Moretto; Emma Openshaw; Peter Cobb; Joshua Dedman (Olsberg·SPI)
Patrick Bradley (Station12)
Alice Fearn; Daniel Hodgkin; Luke Taylor (Verian)

The Foresight Board:

Prof Dave Bull; Prof Darren Cosker; Sarah Ellis; Steve Jelley; Gaby Jenks; Prof Greg Maguire; Prof Alex McDowell; Prof Sylvia Pan; Dr Romana Ramzan; Bill Thompson; Dr Lincoln Wallen; Nell Whitley; Deborah Williams OBE

The Lab is administrated by Petra Lindnerova and Tom Steer.



Introduction

AI presents multifaceted challenges for copyright regulation because AI model training – based on a process known as *text and data mining* (TDM) – relies on access to large amounts of data often containing copyright materials. The AI training process might infringe the copyright owners' rights when the low level of transparency of AI models obstructs right holders from detecting unauthorised use of their materials. Moreover, securing permissions from all copyright owners might not be feasible due to the large amount of data needed for AI training. There is an ongoing debate among scholars, policymakers and rights holders about finding a balance between technological development and encouraging creative expression. This literature review captures this debate around AI and copyright, focusing on academic literature, blog posts and grey literature. Coinciding with the second government consultation on AI and copyright launched in 2024 (HM Government, 2024), the literature review looks at alternative approaches from which the UK might learn. Because of the extent of the current policy debate, there has been a prolific amount of grey literature produced recently on this issue.

This literature review is structured as follows:

- Section 1 provides a context of copyright protection and TDM exceptions adopted by different jurisdictions in the UK, EU, Japan and the US.
- Section 2 explores the arguments put forward by creative professionals regarding the use of their works in AI training and development.
- Section 3 observes the literature exploring alternative frameworks and solutions to the copyright issues.

1. Copyright legislation and TDM exceptions in the UK and beyond

There is a considerable amount of literature discussing the current copyright legislation existing in the UK and worldwide in relation to AI development (Senftleben et al., 2022; Quintais, 2023; Sag, 2011; Thongmeensuk, 2024; Kop, 2019; Kariyawasam, 2021; Iglesias et al., 2021; Bonadio & McDonagh, 2020; Dermawan, 2023; Rosati, 2024; Rosati, 2018; Quintais, 2025). The copyright legislation implies that any reproduction, extraction or reuse of protected data requires permissions from rights-holders unless exceptions are applied. Various jurisdictions have adopted TDM exceptions to foster the development of AI, as discussed in more detail in this section. Copyright limitations and exceptions are limits to the exclusive rights of authors to control the use of their works. There has been a public debate about different legal frameworks of TDM exceptions and their impacts on AI and creative industries. In particular, views vary regarding whether the copyright-protected materials used for AI training infringe the copyright law or fall under TDM exceptions adopted by different legislations.

1.1 UK

The current UK copyright legislation aims at fostering technological progress and innovation and allows using TDM for research with non-commercial purposes: the TDM exception was adopted in 2014 and is active under the Copyright, Designs and Patents Act 1988 (CDPA). In 2021, the UK implemented its National AI Strategy that aimed to 'invest and plan for the long-term needs of the AI ecosystem', 'support the transition to an AI-enabled economy, capturing the benefits of innovation in the UK, and ensuring AI benefits all sectors and regions', and 'ensure the UK gets the national and international governance of AI technologies right' (HM Government, 2021: 7). In 2022, the government launched the first consultation and published its outcome which proposed a broader copyright exception allowing the use of copyrighted content without a licence for any purposes. After receiving negative feedback from the creative sector, the government decided not to implement the proposed exception (Thongmeensuk, 2024; Love et al., 2023). In 2024, the government launched the second consultation proposing an opt-out mechanism for copyright holders who wish to reserve their rights (HM Government, 2024).

1.2 EU

Other legislations, including the EU, have included TDM exceptions in their copyright laws with the aim of facilitating technological progress and innovation. In 2019, the EU implemented a Directive on Copyright in the Digital Single Market (EU CDSMD) that had two TDM exceptions. TDM is defined in Article 2(2) of CDSMD as 'any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations.' The first exception allows TDM for scientific research purposes by research organizations and cultural heritage institutions. Some critiques of this approach note that there is no clear definition of scientific research, which might result in challenges in interpretations (Kollár, 2019). The second exception allows the reproduction and extraction of lawfully accessed materials for TDM purposes, including commercial ones, but copyright holders have the right to monetize their intellectual property by requesting fees or to opt-out to reserve their rights. Adopted in 2024 the EU AI Act offered further protection of copyright holders' interests (Quintais, 2025). The Act stated that the use of copyrighted materials in AI training 'requires the authorisation of the rights holder concerned unless relevant copyright exceptions and limitations apply'. AI developers must obtain the permission of copyright holders before using their materials in the training and development of AI models unless the above-mentioned TDM exceptions of CDSMD are employed. Moreover, GPAI (General-purpose AI) model providers are obliged to be transparent about the content used for the training and development of AI models.

1.3 Japan

Japanese copyright legislation includes a range of exceptions and limitations, and therefore resembles, to a certain degree, the European model (Senftleben et al., 2022). The 'IP Strategic Programme 2016' proposed flexible copyright exceptions to facilitate AI development, and in 2019, it became an amendment to the

Japanese Copyright Act (Thongmeensuk, 2024: 283). It allowed the use of copyright-protected works for TDM for commercial and non-commercial non-enjoyment purposes without securing the permissions of copyright holders. The Japanese legislation distinguished between the enjoyment and non-enjoyment of copyrighted works. Enjoyment of works refers to appreciating works by humans, for example, by listening to music, watching films or reading books, and the copyright law protects it by guaranteeing remuneration to creators. Non-enjoyment, on the other hand, implies that works are used for other purposes, such as analysis or data extraction. It does not entail copyright protection or harm copyright holders' interests (Ueno, 2020; Thongmeensuk, 2024). Thongmeensuk (2024: 284) argues that the Japanese approach provides an auspicious environment for AI development but overlooks the interests of copyright holders resulting in negative impacts on creative industries.

1.4 USA

US regulation pertaining to generative AI training is based on the concept of fair use of copyrighted works that do not infringe the copyright. TDM is regarded as a fair and transformative use of existing copyrighted materials, and therefore it is permitted without licensing or providing compensation to right holders (Sag, 2011). The fair use of copyrighted works is determined by the following factors: (1) 'the purpose and character of the use, including whether such use is of a commercial nature or is for non-profit educational purposes'; (2) 'the nature of the copyrighted work'; (3) 'the amount and substantiality of the portion used in relation to the copyrighted work as a whole'; (4) 'the effect of the use upon the potential market for or value of the copyrighted work' (U.S. Copyright Act, art.107). Because developers do not have to meet all criteria at once, this scheme allows flexibility in deciding whether a case is a fair use. However, the purpose and market effects are considered the most important factors for determining fair use cases (Samuelson, 2023). Some researchers argued that the fair use principle fosters progress in AI development (Cox, 2015; Thongmeensuk, 2024).

2. TDM and Creative Industries impact

The research literature explores the impacts AI and TDM exceptions are likely to have on the creative sector (Frosio, 2023; Vimpari et al., 2023; Watiktinnakorn et al., 2023; Mirowski et al., 2022). Vimpari et al. (2023) analysed the perceptions and attitudes of professionals towards AI models, specifically text-to-image generation (TTIG) models. Conducted between August and October 2022, their study explored press publications containing views of professional creatives on TTIG. Among the 11 topics identified, copyright was a central theme. Creative professionals expressed a concern that their works were used in the training and development of TTIG without their permission. They explained that images created by those models might copy their ideas and style without giving artists any credit. They also raised a question about remuneration for their work used in the training and development of TTIG (Heikkilä, 2022; Knight, 2022; Savov, 2022; cited

from Vimpari et al., 2023). Furthermore, creative professionals referred to the issue of ownership of works created with TTIG models (Goldman, 2022; cited from Vimpari et al., 2023). For example, one of the artists expressed their resentment about AI technologies:

AI is the latest and most vicious of these technologies. It basically takes lifetimes of work by artists, without consent, and uses that data as the core ingredient in a new type of pastry that it can sell at a profit with the sole aim of enriching a bunch of yacht owners. – Simon Stålenhag, quoted by Will Knight, Wired (Knight, 2022, cited from Vimpari et al., 2023)

In addition to the press study, Vimpari et al. (2023) conducted qualitative interviews with professionals from the Finnish video game industry to explore their perceptions and attitudes towards AI. Creative professionals articulated similar concerns about copyright issues and remuneration for using their works in AI development. They expressed worries about slowly changing legislation on AI and copyright. Similarly, in a study of the approaches and challenges of creative professionals about AI (Watiktinnakorn et al., 2023), participants argued that there was a need for developing new legal frameworks for rapid technology advancements. In quantitative surveys and qualitative interviews, creative professionals expressed their belief that humans and machines were creative collaborators whose joint efforts should be protected by copyright law. The works generated as a result of this collaboration should be included in copyright law provisions (ibid.).

To communicate their concerns about the infringement of their rights, creative professionals have organised multiple campaigns, petitions and open letters in the UK and beyond (Love et al., 2023; Thongmeensuk, 2024). After launching the first consultation in 2022 that proposed the copyright exceptions, the British government received feedback from the creative sector expressing their discontent regarding the proposed changes (Love et al., 2023; Thongmeensuk, 2024). The objection was that the suggested exception did not provide creatives with an opportunity to license and receive remuneration for their works, therefore they might need to 'exit the UK market or apply paywalls where access to content is currently free' (Love et al., 2023). The second consultation launched in 2024 received similar feedback from the creative sector rejecting the proposed plans (Booth, 2024). The Creative Rights in AI Coalition, a body purporting to be representative across the creative industries, released a statement rejecting the government proposal. The statement said:

We support the Government's mission for long-term, secure growth in the creative and tech sectors. We are eager to see the development of a vibrant licensing market and support the sectors which rely on us for their future prosperity, but we can only do so with a robust copyright framework which preserves our exclusive rights to control our works and thereby act as a safeguard against misuse. (Creative Rights in AI Coalition, 2025)

The Creative Rights in AI Coalition stated that the existing copyright law should be 'respected and enforceable', requiring AI companies to seek permissions, agree on licences and offer remuneration to creative professionals. AI companies must implement transparency measures allowing copyright holders to be informed about how their works were used by AI models. Additionally, a petition signed by more than 40,000 creative professionals stated the following: 'The unlicensed use of creative works for training generative AI is a major, unjust threat to the livelihoods of the people behind those works and must not be permitted' (Statement on AI training, 2025).

There were similar campaigns organised worldwide by creative professionals from various industries petitioning for protecting copyright principles and establishing fair licensing agreements (Lucchi, 2024; Frosio, 2023). For example, more than six million creative professionals signed an open letter, 'Global Creators and Performers Demand Creative Rights in AI Proliferation – An Open Letter to policymakers on Artificial Intelligence' organised by the International Confederation of Societies of Authors and Composers (2023). Their demands included protecting their rights against exploitation by AI systems, implementing fair licensing agreements, ensuring transparency, proper recognition and credit. Moreover, they objected to implementing exceptions for TDM that did not guarantee effective opt-out schemes. The artists demanded establishing effective international legislation ensuring ethical and responsible AI use and development. Another example included an open letter from the Author's Guild (2023) signed by more than 15,000 writers to Generative AI leaders such as OpenAI, Alphabet, Meta, Stability AI, IBM, and Microsoft to obtain consent and compensate writers fairly for the use of their works in AI development.

3. Recommendations, frameworks and suggested solutions to copyright issues

To respond to the challenges posed by copyright holders and AI developers, researchers suggested various frameworks, codes of best practice, and alterations to the development of AI models. They aimed to create an environment which would incentivise content creators and stimulate AI development. This section reviews proposed ideas regarding obtaining copyright holders' permissions, providing fair remuneration, establishing a new AI right, facilitating an open data sharing community, adopting ethical standards, ensuring transparency in data handling and use, and best practice recommendations to stakeholders, including AI companies, copyright holders and users.

3.1 Obtaining the copyright holders' permissions and agreements ex-ante

It is a widely accepted approach among researchers that TDM involves copying and reusing data and therefore AI companies must secure permissions and agreements with copyright owners before using their works for the training and development of AI (Appel et al., 2023; Lucchi, 2024; Senftleben et al., 2022;

Silver, 2025; Geiger & Iaia, 2024; Sobel, 2017). Appel et al. (2023) argued that it is the primary responsibility of AI developers to secure IP to the works before using it in their AI model training. Rather than expecting copyright holders to opt out ex-post, AI companies must obtain the copyright holders' opt-in permission ex-ante (ibid.). Similarly, Lucchi (2024) suggested making data-sharing agreements to protect copyright holders' rights. Those agreements would contain necessary permissions and licences for using works in AI training. Moreover, agreements would present a legal framework establishing a scope of data use, access and administration of copyright-protected works in AI models. Lucchi (2024) proposed including information about AI models, algorithms, timing, duration of using protected data and any limitations. The agreements also would include information about fair compensation to copyright holders and the method for calculating compensation (ibid.). Senftleben et al. (2022: 73) argued that licensing agreements would enhance cooperation between AI companies and creative sectors and stimulate progress towards the use of protocols and safeguards that would benefit both parties.

However, some researchers argued that AI developers should be able to use datasets for training purposes regardless of copyright protection, on account of the existing copyright exceptions as well as there being a challenging and time-consuming process of securing IP for works required for AI training (Dermawan, 2023; Thongmeensuk, 2024; Lemley & Casey, 2020). Lemley & Casey (2020) argued that training data usually had a large amount of works and owners, whose permissions would be implausible to secure for AI developers. They claimed that the main aim of AI training was learning underlying facts, ideas and functions that were not under the scope of copyright laws (ibid.; also, Dermawan, 2023). To facilitate AI development and safeguard the interests of copyright holders, Thongmeensuk (2024) proposed a hybrid model of TDM exceptions based on the analysis of the existing copyright regulations worldwide. The hybrid model distinguished between two categories of TDM exceptions – for non-commercial and commercial purposes. The non-commercial TDM would be a mandatory exception serving public interests of research and scientific progress, and it could be utilised by both public and private organisations. The commercial TDM would imply two essential conditions: it would guarantee transparency provision for the use of data in AI training and an opt-out clause for copyright holders. Thongmeensuk (2024) suggested using the EU approach for the opt-out mechanism which would allow copyright holders to opt out from commercial AI training. Thongmeensuk (2024) claimed that this hybrid model would alleviate the AI developers' burden of securing permissions from copyright holders and respect and acknowledge the interests of copyright holders. However, if the former seems plausible, the latter is unlikely, given the unequal distribution of resources between AI companies and individual copyright holders. Proposing the opt-out mechanism would further benefit AI companies who are already in a privileged position compared to individual copyright holders.

3.2 Fair remuneration

There is a considerable amount of literature proposing implementing fair remuneration to copyright holders via revenue sharing or royalty payments (Appel et al., 2023; Lucchi, 2024; Samuelson, 2023; Silver, 2025; Wang et al., 2024; Geiger & Iaia, 2024; Geiger, 2024; Geiger et al., 2024; Kop, 2020b; Quintais, 2025; Sobel, 2017; Senftleben, 2023). Silver (2025) argued that licensing should become a 'friction-free commercial service', which would make AI tools a 'valuable revenue stream' for copyright holders. Geiger et al. (2024) suggested introducing 'limitation-based remuneration rights', implying that authors whose works were being used for AI training had non-exclusive rights. The remuneration rights would be activated every time when works were used by others. According to Geiger et al. (2024), this scheme would reward creators fostering both creativity and innovation. Wang et al. (2024) suggested a solution to the copyright issues from an economic perspective fostering 'a mutually beneficial revenue-sharing agreement between AI developers and copyright owners' (p.2). Their framework would pay royalties to copyright holders that are proportional to their contributions to the AI-generated content. Using techniques from cooperative game theory in economics, particularly the theory of Shapley value, the researchers proposed quantitative metrics for contributions and associated fair fees. The metrics allow transparency and interpretability in the distribution of royalties. They also advised on creating a platform that would offer high-quality training data to AI developers. The research (ibid.) demonstrated the effectiveness of the framework by conducting experiments assessing its ability to fairly distribute royalties among various copyright holders, including creative art, logo design, and language generation. Similarly, Lucchi (2024) suggested establishing a revenue-sharing agreement, according to which copyright holders would receive a portion of the revenue generated by AI companies and it would be relational to their contribution to the training dataset. Royalty-based remuneration, on the other hand, would provide a fixed amount of fee for the use of protected works by AI models (ibid.).

Senftleben (2023) and Frosio (2023) proposed an alternative approach to remuneration and recommended the introduction of an AI levy system. The levy system would oblige AI providers to make general payments, rather than payments based on AI training activities. This system would contribute revenue from AI companies towards social and cultural funds with collective management of those funds (Frosio, 2023). Senftleben (2023: 1537) argued that the funds could be used to provide support to artistic projects and training opportunities. According to Senftleben (2023) and Frosio (2023), this system would support creative work and simultaneously strengthen AI sector development.

3.3 New AI right

Moving a step forward, Silver (2025) suggested creating a new AI IP right that would be separate from the existing copyright law, specifically the Copyright Designs and Patents Act 1988 in the UK. This new IP right would demonstrate that AI training is not copying in the traditional copyright law sense, but rather 'the creation of derivative works whose provenance may be traced' (ibid.). The AI IP right would allow or forbid

the use of protected materials for AI training, and it would be the responsibility of AI companies to maintain transparency regarding data sources and obtain permissions and licences (ibid.). The AI IP right would require a new global rights body – Collective Management Organisation (CMO) – that would enforce the new law and be sponsored by the tech platforms. CMO would need to have regulatory power globally to authorise permissions, licences and remuneration.

Similarly, Kop (2020a) proposed introducing into legislation a new right to process data for machine learning purposes. Rejecting ideas about private property on data, Kop argued that data is a common good that should be governed by common property rules. However, there is a lack of legislation on the use of data for machine learning purposes. Kop proposed to implement a new legal basis for access, sharing and re-use of machine learning data aiming to fulfil public interests. The scope of the right incorporates all types of data, and industry-specific regulations could clarify rules further. According to Kop (2020a), this new vision of data governing would foster innovation and technological progress.

3.4 Open-source training datasets

To facilitate AI development, some studies suggested creating open-source datasets or repositories for the training of AI models (Senftleben et al., 2022; Lemley & McCreary, 2021; Lucchi, 2024; Kop, 2019; Kop, 2020a; Dermawan, 2023). As opposed to the current ownership of data by large corporations, open datasets would facilitate collaboration, equity and transparency in AI development. Kop (2019) suggested that creating a 'global open data sharing community' with comprehensive regulation policies would promote transparency and facilitate economic growth. Kop (2020a) wrote: 'No exclusive control over training data should be allowed. It should be free as the air we breathe, for common use.' (p.13). Lucchi (2024) argued that free and open training data would enable AI development for the collective benefit of society, including marginalised communities. These open datasets would play a role of 'public participatory good' which is an outcome of collective efforts (ibid.). The repositories would be centralised platforms simplifying the process of obtaining permissions and licences from copyright holders (Senftleben et al., 2022; Lucchi, 2024). Likewise, Senftleben et al. (2022) argued that an improved and harmonised data infrastructure would benefit creative industries as much as AI companies. Similarly, Lemley & McCreary (2021) proposed that specific AI training databases should become available to all AI systems and AI companies should share access with other companies to their databases. Moreover, it is imperative that appropriate legislation is created to foster the creation and maintenance of those open datasets. Lucchi (2024) also argued that legislators should encourage companies to share their datasets, by, for example, providing grant funding to organisations that contribute to open repositories. The new regulations should also include guidance on ethical data management, safeguarding, and preventing harm and discrimination.

Some researchers described risks posed by open-source data (Kop, 2020a; Zuidervijk & Janssen, 2014). Open-source data can result in misinterpretation and misuse of data, which can be avoided by thorough

investigations. Open datasets might result in various security and privacy challenges for stakeholders whose interests would be at risk. One of the solutions to this problem might be excluding a subject at risk from datasets or incorporating restricted data-sharing arrangements (Kop, 2020a: 16).

3.5 Ethical standards, transparency and AI reporting

The research suggested that AI companies should adopt ethical guidance and standards for AI training and development (Appel et al., 2023; Geiger & Iaia, 2024; Lucchi, 2024; Senftleben et al., 2022; Silver, 2025). The ethical standards would regulate all processes involved in AI development, such as data collection, consent, algorithms and training (Lucchi, 2024). The transparency mechanisms would require companies to disclose information about the protected content used in AI training and identify copyright holders (Geiger & Iaia, 2024; Silver, 2025; Thongmeensuk, 2024). Transparency would require AI companies to be responsible and accountable for respecting the interests of copyright holders (Thongmeensuk, 2024). Silver (2025) added that making visible the sources used in AI would also contribute to the fight against misinformation. AI companies were advised to maintain transparency by fostering AI reporting regarding various settings, platforms used for developing content, metadata, prompts, etc. (Appel et al., 2023). However, thorough guidance is needed for the practical implementation of transparency provisions, such as defining what would be considered suitably detailed information about copyright-protected materials used in AI training (Thongmeensuk, 2024).

Some researchers argued that the transparency requirement might be unfeasible to achieve for AI companies (Geiger & Iaia, 2024; Quintais, 2023; Thongmeensuk, 2024). Quintais (2023) maintained that it would be unrealistic to ask AI companies to provide information about all copyrighted works used in AI training or copyright ownership claims. The challenges result from various factors such as a low threshold of originality, the complexity of copyright, which might be fragmented across multiple ownerships and legislations, and insufficient information on copyright works due to the absence of a mandatory registration (ibid.). Addressing these issues would facilitate transparency in AI development. Thongmeensuk (2024) suggested using AI technology to solve the problems: instead of asking AI developers to provide a list of copyright-protected works, AI could scan and identify those materials within data to comply with the transparency requirement.

3.6 Best practices for AI companies and technical mitigations

Sag (2023) proposed Best Practices for Copyright Safety for Generative AI noting that there was a need to develop guidance for the safe use of copyrighted works by AI companies. Sag argued that it was imperative to use copyrighted works for the training of AI because the public domain and open licence materials currently were very limited, and therefore their use for the training of AI models would be an unfavourable solution (Levendowski, 2018, cited from Sag, 2023). The study (Sag, 2023) included various technical suggestions for training LLMs ex ante, including removing duplicates of copyrighted works from the training

data, instructing LLMs to learn abstractions rather than details, and using reinforcement learning through human feedback to address copyright issues. Moreover, to avoid copyright infringement, developers of LLMs were advised to create strategies for filtering model outputs and to restrict access to those LLMs that presented risks for copyright infringement. AI developers were recommended to be transparent and keep a detailed record of copyrighted works used as training data (ibid.).

Similarly, several researchers have suggested various technical mitigations to reduce the copyright infringement possibility (Chu et al., 2023; Kirchenbauer et al., 2023; Shan et al., 2023; Henderson et al., 2023). Henderson et al. (2023) proposed multiple solutions including 'data and output filtering'; 'instance attribution'; 'differentially private training'; and 'fair use alignment via learning from human feedback' (p.26). Chu et al. (2023) offered a method of softmax regression stopping the function from generating copyright data. Their model modified AI training objectives to avoid creating outputs resembling copyright-protected data. To prevent copyright infringement harm, Kirchenbauer et al. (2023) suggested a watermarking framework that would insert invisible signals into generated outputs, which would be evident for algorithms only. Shan et al. (2023) developed a tool allowing creative professionals to embed a 'style cloak' into their art to avoid AI mimicking their artistic style. The tool would incorporate perturbations to images, which would deceive generative AI attempting to imitate artistic styles.

3.7 Searchable portal for copyright holders

To ensure the protection of the interests of right holders, some studies have suggested creating searchable portals for detecting the presence of copyright-protected works (Thongmeensuk, 2024; Appel et al., 2023). Those portals would empower copyright holders seeking to protect their IP. Similarly, Appel et al. (2023) advised content creators to make efforts to monitor with automated tools if their works were used in datasets to avoid copyright infringement harm. Some websites, such as 'Have I Been Trained?', offer a service detecting the use of copyright materials in existing AI models (Kamps & Wiggers, 2022). Additionally, Dermawan (2023) suggested using a robot exclusion standard as a warning sign that TDM was not permissible on a website. The machine-readable file would restrict TDM activities and safeguard the interests of copyright holders.

3.8 Responsible use of AI

The research literature also considered the responsible use of AI an important point for the copyright and AI discussion. Appel et al. (2023) provided recommendations for best practices for all stakeholders involved in working with AI tools, including their users. AI users were advised to review the terms of service and privacy policies of AI tools to ensure that their training data was licensed from content creators (ibid.). Also, businesses were recommended to include protections in their contracts and require terms of services from AI companies (ibid.). Frosio (2023) argued that responsibility for violation of copyright infringement predominantly relied on AI users who were involved in generating illegal materials. Frosio (2023) claimed that AI platforms were 'neutral or dual-use technology' (p.20) that could be used for legitimate and unlawful purposes.

4. Conclusion

AI development is highly dependent on the availability of large amounts of data, for millions of texts, images, and videos are required for the training of AI models. When datasets contain protected works, there is a risk that AI models might infringe copyright law and have a significant impact on the creative sector. This literature review has captured the scholarly and policy debates about the issues of access to data and copyright legislation. To balance the interests of right holders and AI developers, researchers proposed multiple legislative reforms to establish an ethical and responsible AI environment, including introducing licencing, fair remuneration, open-source training datasets, ethical standards, and transparency requirements. Scholars also suggested harmonising copyright legislation to encourage international collaborations in AI development.

Suggestion citation:

- Tarnovskaya, E. and Hitchen, G. (2025). *Copyright and Artificial Intelligence - Literature Review*. doi: 10.5281/zenodo.15000974

References

- Appel G, Neelbauer J & Schweidel D (2023) *Generative AI Has an Intellectual Property Problem*. Available at <https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>
- Bonadio E & McDonagh L (2020) 'Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity'. *Intellectual Property Quarterly*, 2, pp. 112-137. <https://ssrn.com/abstract=3617197>
- Center for Artistic Inquiry and Reporting (2023) *Restrict AI Illustration from Publishing: An Open Letter*. Available at <https://artisticinquiry.org/AI-Open-Letter>
- Chu T, Song Z & Yang C (2023) *How to protect copyright data in optimization of large language models?* <https://doi.org/10.48550/arXiv.2308.12247>
- Cox K (2015) *Text and Data Mining and Fair Use in the United States*. Association of Research Libraries. Available at: <https://www.arl.org/resources/text-and-data-mining-and-fair-use/>
- Creative Rights in AI Coalition (2025) *Statement on creative rights in AI*. Available at: <https://www.creativerightsinai.co.uk/>
- Cuenco M (2023) *We must declare jihad against AI*. Compact. Available at <https://www.compactmag.com/article/we-must-declare-jihad-against-a-i/>
- Frosio G (2023) *Should we ban generative AI, incentivise it or make it a medium for inclusive creativity?* In: Bonadio E, Sganga C (eds) *A research agenda for EU copyright law*. Edward Elgar, Cheltenham (forthcoming). <https://ssrn.com/abstract=4527461>.
- Geiger C. (2024). *When the Robots (Try to) Take Over: Of Artificial Intelligence, Authors, Creativity and Copyright Protection*. In: Thouvenin, F., Peukert, A., Jaeger, T., Geiger, C. (eds) *Kreation Innovation Märkte - Creation Innovation Markets*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-68599-0_6
- Geiger C & Iaia V (2024) 'The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI'. *Computer Law & Security Review*, 52. <https://doi.org/10.1016/j.clsr.2023.105925>
- Goldman S (2022) *Will openai's dall-e 2 kill creative careers?* Available at <https://venturebeat.com/ai/openai-will-dall-e-2-kill-creative-careers/>
- Heikkilä M (2022) *This artist is dominating AI-generated art and he's not happy about it*. Available at <https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>
- Henderson P et al. (2023) *Foundation Models and Fair Use*. arxiv.org. Available at <https://arxiv.org/abs/2303.15715>
- HM Government (2021) *National AI Strategy*. Available at: https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf
- HM Government (2024) *Copyright and AI: Consultation*. Available at: <https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence/copyright-and-artificial-intelligence>
- Iglesias M., Shamuilia S. and Anderberg A., (2021) *Intellectual Property and Artificial Intelligence – A literature review*. European Commission: Joint Research Centre Publications Office. <https://data.europa.eu/doi/10.2760/8600>
- International Confederation of Societies of Authors and Composers (2023) *Global Creators and Performers Demand Creative Rights in AI Proliferation – An Open Letter to policy makers on Artificial Intelligence*. Available at: <https://www.cisac.org/Newsroom/articles/global-creators-and-performers-demand-creative-rights-ai-proliferation>

- Kamps H & Wiggers K (2022) *This Site Tells You If Photos of You Were Used to Train the AI*. TechCrunch. Available at: <https://techcrunch.com/2022/09/21/who-fed-the-ai/>
- Kariyawasam K (2021) Artificial intelligence and challenges for copyright law. *International Journal of Law and Information Technology*, Volume 28, Issue 4
- Kirchenbauer J, Geiping J, Wen Y, Katz J, Miers I, and Goldstein T (2023) *A watermark for large language models*. arXiv.org. <https://doi.org/10.48550/arXiv.2301.10226>
- Klein N (2023) *AI machines aren't 'hallucinating.' But their makers are*. The Guardian. Available at: <https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein>
- Knight W (2022) *Algorithms can now mimic any artist. Some artists hate it*. Available at <https://www.wired.com/story/artistsrage-against-machines-that-mimic-their-work/>
- Kollár P. (2021) *Mind if I Mine? A Study on the Justification and Sufficiency of Text and Data Mining Exceptions in the European Union*. Available at <http://dx.doi.org/10.2139/ssrn.3960570>
- Kop M. (2019). *AI & Intellectual Property: Towards an Articulated Public Domain*. <https://ssrn.com/abstract=3409715>
- Kop M. (2020a) *The Right to Process Data for Machine Learning Purposes in the EU*. *Harvard Journal of Law & Technology (JOLT)* Forthcoming, Available at SSRN: <https://ssrn.com/abstract=3653537>.
- Kop M. (2020b) *Machine Learning & EU Data Sharing Practices*. *Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 1/2020*, Available at SSRN: <https://ssrn.com/abstract=3409712>
- Lemley M & Casey B (2020) *Fair Learning*. Available at SSRN: <https://ssrn.com/abstract=3528447>
- Lemley M. and McCreary, A. (2019) *Exit Strategy*. *Stanford Law and Economics Olin Working Paper #542*, Available at SSRN: <https://ssrn.com/abstract=3506919>
- Levendowski A. (2017). *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*. Available at SSRN: <https://ssrn.com/abstract=3024938>
- Love J et al. (2023) *Entertainment and Media Guide to AI: Text and Data Mining around the Globe*. Reed Smith LLP. Available at <https://www.reedsmith.com/en/perspectives/ai-in-entertainment-and-media>
- Madigan K (2024) *Top Takeaways from Order in the Andersen v. Stability AI Copyright Case*. Copyright Alliance. Available at: <https://copyrightalliance.org/andersen-v-stability-ai-copyright-case/#:~:text=In%20early%202023%2C%20visual%20artists,use%20of%20the%20plaintiffs'%20works>
- Mirowski P, Mathewson K, Pittman J, and Evans R (2022) *Co-writing Screenplays and Theatre Scripts with Language Models: An Evaluation by Industry Professionals*. <http://arxiv.org/abs/2209.14958>
- Quintais J (2023) *Generative AI, Copyright and the AI Act*. Kluwer Copyright Blog, available at: <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act>
- Quintais J (2025) *Generative AI, copyright and the AI Act*. *Computer Law & Security Review* Volume 56. <https://doi.org/10.1016/j.clsr.2025.106107>.
- Rosati E (2018) *The exception for text and data mining (TDM) in the proposed Directive on Copyright in the Digital Single Market – Technical aspects*. European Parliament. Available at <https://data.europa.eu/doi/10.2861/480649>
- Rosati E (2024) *Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law*. *European Journal of Risk Regulation, First View*, pp. 1 - 25 DOI: <https://doi.org/10.1017/err.2024.72>

- Sag M (2023) *Copyright safety for generative ai*. *Houston Law Review*, 61(2):295–347. Available at: <https://houstonlawreview.org/article/92126-copyright-safety-for-generative-ai>
- Sag M (2009) *Copyright and Copy-Reliant Technology*. *Northwestern University Law Review*, Vol. 103, 2009, The DePaul University College of Law, Technology, Law & Culture Research Series Paper No. 09-001, Available at SSRN: <https://ssrn.com/abstract=1257086>
- Samuelson P (2023) *Generative ai meets copyright*. *Science*, 381(6654):158–161. <https://www.science.org/doi/abs/10.1126/science.adi0656>
- Savov V (2022) *Dall-E and Midjourney and the future of AI Art Generation*. Available at <https://www.bloomberg.com/news/newsletters/2022-08-16/dall-e-and-midjourney-and-the-future-of-ai-art-generation>
- Senftleben M (2023) *Generative AI and Author Remuneration*. *International Review of Intellectual Property and Competition Law* 54 (2023), pp. 1535-1560, Available at SSRN: <https://ssrn.com/abstract=4478370> or <http://dx.doi.org/10.2139/ssrn.4478370>
- Senftleben M, Margoni T, et al. (2022) *Ensuring the Visibility and Accessibility of European Creative Content on the World Market - The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive* 13. *JIPITEC* 67 para 1. <https://www.jipitec.eu/jipitec/article/view/345>
- Shan S, Cryan J, Wenger E, Zheng H, Hanocka R, and Zhao B (2023) *Glaze: Protecting artists from style mimicry by text-to-image models*. arXiv. <https://doi.org/10.48550/arXiv.2302.04222>
- Silver J (2025) *Copyright and AI - a new AI Intellectual Property Right for composers, authors and artists*. Creative Industries Policy and Evidence Centre. https://pec.ac.uk/blog_entries/copyright-and-ai-a-new-ai-intellectual-property-right-for-composers-authors-and-artists/
- Sobel B (2017) *Artificial Intelligence's Fair Use Crisis*. *Columbia Journal of Law & Arts*. <https://doi.org/10.7916/jla.v4i1.2036>
- *Statement on AI training* (2025) Available at: <https://www.aitrainingstatement.org/>
- Ueno T (2020) *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes—Recent Amendment in Japan and Its Implication*. 70(2) *GRURI* 148. <https://doi.org/10.1093/grurint/ikaa184>
- Wang J T, Deng Z, Chiba-Okabe, H, Barak B, Su W J (2024) *An Economic Solution to Copyright Challenges of Generative AI*. arXiv.org. <https://doi.org/10.48550/arXiv.2404.13964>
- Watiktinnakorn C, Seesai J & Kerdvibulvech C (2023) *Blurring the lines: how AI is redefining artistic ownership and copyright*. *Discover Artificial Intelligence* Volume 3, article number 37. <https://doi.org/10.1007/s44163-023-00088-y>



CoSTAR Foresight Lab
Goldsmiths, University of London
New Cross
London SE14 6NW

costarforesightlab@gold.ac.uk
costarnetwork.co.uk/labs/foresightlab



THE UNIVERSITY
of EDINBURGH



Loughborough
University

Goldsmiths
UNIVERSITY OF LONDON