ALL AIS ON THE EU AI ACT'S COPYRIGHT PROVISIONS: NAVIGATING A NEW LEGAL FRONTIER

by DR. BARRY SCANNELL*

This article examines the EU AI Act's implications for copyright law, focusing on its treatment of General Purpose AI models and text and data mining. It analyses key provisions, including transparency obligations, copyright reservations, and extraterritoriality concerns under Article 53 and Recital 106. By exploring the Hamburg Court's recent LAION decision and considering potential intermediate solutions, this article highlights regulatory challenges and opportunities for AI developers and copyright holders.

INTRODUCTION

The intersection of artificial intelligence (AI) with intellectual property law—especially copyright—has become increasingly complex. At the forefront of this issue is the European Union's AI Act, which not only seeks to regulate the development and deployment of AI systems within the EU but also establishes

* Dr. Barry Scannell is a globally recognized expert in AI law and governance. A partner in one of Ireland's top law firms, William Fry, Barry has been a driving force in shaping legal strategies for some of the world's leading technology companies. Specializing in AI and intellectual property law, he advises clients on Irish and EU legal frameworks, helping them navigate the complexities of regulatory compliance. In January 2024, Barry was appointed by the Irish Government to Ireland's AI Advisory Council. His PhD is on AI and Copyright. Barry also sits on the Law Society of Ireland's IP and Data Protection Committee.

guidelines that could have far-reaching implications for copyright law globally.[1] The AI Act addresses critical questions surrounding the use of copyrighted materials in AI training, with particular attention to General Purpose AI (GPAI) models, and introduces obligations that AI developers and providers must follow to ensure compliance with EU copyright law.[2]

This short article explores the key aspects of the EU AI Act that relate to copyright, the challenges it poses to AI developers, and the broader impact it may have on the global AI ecosystem.

I. RECITALS AND ARTICLES IN EU REGULATIONS

Recitals and articles in EU regulations differ in their purpose and function. *Recitals* are placed at the beginning of the regulation and provide the background, reasoning, and objectives behind the legislation.[3] They explain why the regulation was introduced and what it aims to achieve but are not legally binding.

In contrast, *articles* are the legally enforceable parts of the regulation. They set out the specific rules, obligations, and procedures that must be followed by Member States, companies, or individuals. While recitals help interpret the articles, they do not have the same legal force as the articles themselves.[4] This distinction is important as there appears to be some issues between the recitals and articles in the AI Act relating to copyright, given that Recital 106 as we will see below, appears to say that Article 53 of the AI Act relating to copyright has extraterritorial effect, but this extraterritoriality is not referred to in the Article itself.[5]

II. GENERAL PURPOSE AI MODELS AND COPYRIGHT

At the heart of the EU AI Act's provisions on copyright is the regulation of GPAI models—those that are capable of performing a wide range of tasks without being specifically designed for any one function.[6] These models, which often rely

---

[1] Regulation 2024/1689, of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828, 2024 O.J. (L) 1 [hereinafter Artificial Intelligence Act].

[2] *Id.* at 27-28, 84.

[3] *See Recital*, BLACK'S LAW DICTIONARY (12th ed. 2024).

[4] *Recitals (EU)*, THOMAS REUTERS GLOSSARY (2024) ("Recitals to EU laws are not in themselves legally binding in the same way that the operative provisions are. However, where an EU law is ambiguous, the recitals can be important in interpreting the ambiguous provision.").

[5] *Infra* note _.

[6] *See generally* Artificial Intelligence Act, *supra* note 1 (distinguishing "general-purpose" AI systems over two hundred times throughout the Regulation).

on large-scale datasets for training, represent both innovation and challenges for the protection of copyrighted works.[7]

Recital 105 of the AI Act makes clear that large generative AI models, such as those used to generate text, images, or other creative content, raise concerns about how these systems access and use copyrighted material.[8] The development of such models requires extensive use of data, including text and images that may be protected by copyright and related rights. The use of text and data mining (TDM) techniques to retrieve and analyse these materials is central to the training process.[9] However, such use often requires the authorisation of the rightsholders, unless relevant exceptions and limitations to copyright apply, as outlined in the EU's Copyright Directive (Directive (EU) 2019/790).[10]

III. THE COPYRIGHT DIRECTIVE AND TEXT AND DATA MINING

The 2019 Copyright Directive introduced specific exceptions for text and data mining in Articles 3 and 4.[11] Article 3 allows for an exception to copyright for reproductions and extractions of works for scientific research purposes by research organisations and cultural heritage institutions.[12] However, this exception is narrowly tailored to non-commercial activities and requires that rightsholders be allowed to apply security measures to protect their works from unauthorised use.[13]

Article 4 extends the exception for text and data mining to a broader scope, allowing reproductions and extractions of lawfully accessible works for commercial purposes, provided that the rightsholders have not expressly reserved their rights in an appropriate manner.[14] In practical terms, this means that if a rightsholder chooses to reserve their rights over a particular work, such as by using a machine-readable format like a robots.txt file, AI developers must seek permission to use that content for text and data mining.

---

[7] *See General Purpose AI and the AI Act*, FUTURE OF LIFE INSTITUTE 1, 3 (May 2022), https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf.

[8] Artificial Intelligence Act, *supra* note 1, at 27 ("General-purpose AI models, in particular large generative AI models, capable of generating text, images, and other content, present unique innovation opportunities but also challenges to artists, authors, and other creators and the way their creative content is created, distributed, used and consumed.").

[9] *Id.*

[10] *Id.*; Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC O.J. (L 130) 92, 113 [hereinafter Directive 2019/790].

[11] Directive 2019/790, *supra* note 10, at 113-14.

[12] *Id.*

[13] *Id.*

[14] *Id.*

IV. RECITAL 106 AND THE EXTRATERRITORIAL APPLICATION OF EU COPYRIGHT LAW

One of the most controversial aspects of the AI Act is its potential extraterritorial reach, particularly when it comes to copyright compliance. Recital 106 of the Act states that providers placing GPAI models on the Union market must comply with EU copyright law, regardless of where the AI training takes place.[15] This provision has raised concerns about whether AI developers outside the EU will be required to navigate not only their own local copyright laws but also those of the EU if they wish to operate within the bloc.[16]

The intent behind Recital 106 is clear: to ensure that no AI provider gains a competitive advantage by training their models in jurisdictions with more lenient copyright standards and then selling or deploying those models in the EU. This approach aims to create a level playing field for AI providers and to ensure that the EU's high copyright standards are respected, even if the training of the AI occurs outside the Union.

However, as outlined above, recitals are not binding in EU legislation.[17]

V. ARTICLE 53: OBLIGATIONS FOR PROVIDERS OF GENERAL PURPOSE AI MODELS

Article 53 of the AI Act establishes the binding specific obligations for providers of GPAI models.[18] These include the requirement to put in place a policy to comply with EU copyright law, particularly with respect to the reservation of rights expressed by rightsholders under Article 4(3) of the Copyright Directive. Providers must ensure that they have the necessary authorisations to carry out text and data mining on works for which rights have been reserved.[19]

This provision means that AI providers must actively monitor and respect the reservations of rights expressed by copyright holders, and they must employ state-of-the-art technologies to ensure compliance. For companies that operate globally, this could result in significant operational challenges, as they may be required to

---

[15] Artificial Intelligence Act, *supra* note 1, at 28 ("Any provider placing a general-purpose AI model on the Union market should comply with this obligation, regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place.").

[16] *See, e.g.*, Sean Musch et al., *The EU AI Act As Global Artificial Intelligence Regulation* (Aug. 23, 2023), http://dx.doi.org/10.2139/ssrn.4549261.

[17] *Recital, supra* note 4.

[18] Artificial Intelligence Act, *supra* note 1, at 84-85 (Providers shall "draw up and keep up-to-date the technical documentation of the model," "draw up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems," "put in place a policy to comply with Union law on copyright and related rights," and "draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.").

[19] *Id.* at 28.

comply with both EU copyright law and the copyright laws of other jurisdictions. However, unlike Recital 106, there is no mention of extraterritoriality in Article 53, leading to confusion as to the extent of the requirements under the legislation.

It is not clear whether it is simply enough pursuant to Article 53 to have a policy in place which outlines how EU copyright law will be complied with. If a model is trained in jurisdictions like Japan or Singapore which have legal regimes which broadly permit text and data mining of copyright materials, that training does not breach EU copyright law, as EU copyright law does not apply to such training. One interpretation of Article 53(1)(c) of the AI Act could be that a GPAI model provider would only need to have a policy on EU copyright law compliance in place, which would only be relevant when EU copyright law applies. However, that argument is countered by the non-binding Recital 106's "*regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place*" language. This potentially contradictory language between Recital 106 and Article 53 has created unhelpful ambiguity in terms of copyright obligations under the AI Act.

VI. TRANSPARENCY OBLIGATIONS AND COPYRIGHT COMPLIANCE

Another key aspect of the AI Act is its emphasis on transparency.[20] Recital 107 and Article 53 require providers of GPAI models to draw up and make publicly available a summary of the content used to train their models.[21] This summary should include information about the datasets or data collections that were used, while still protecting trade secrets and confidential business information.

The goal of these transparency requirements is to ensure that copyright holders, and other stakeholders with legitimate interests, can easily access information about the use of their works in AI training. This, in turn, will allow rightsholders to exercise their rights and take legal action if necessary to enforce copyright protections.

---

[20] *Id.* at 8 (naming "transparency" as one of the seven non-binding ethical principles for AI).

[21] *Compare id.* at 28 (Providers must "draw up and make publicly available a sufficiently detailed summary of the content used for training the general-purpose AI model" which is "generally comprehensive in its scope" to protect trade secrets and confidential information.) *with id.* at 84 (Providers must "draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.").

VII. THE HAMBURG COURT DECISION: LAION AND THE RESEARCH EXCEPTION

The recent Hamburg court decision regarding LAION, a non-profit research organization, and German photographer Robert Kneschke is a landmark ruling in the evolving legal landscape surrounding text and data mining (TDM) and generative AI.[22] The case offers insight into the intricate balance between copyright protection and the broader use of data for AI training.

The court ruled that LAION could rely on the research exception under Section 60d of the German Copyright Act (UrhG) to justify the inclusion of Kneschke's image in the LAION-5B dataset.[23] This decision underlines the importance of the research exception within European copyright law, particularly as it applies to non-commercial entities. For-profit organisations, on the other hand, are not afforded the same protection, and this ruling raises significant implications for tech companies relying on similar datasets for commercial purposes.

One of the most contentious aspects of the case involved LAION's claim that it could avail itself of the research exception under Article 3 of the Copyright Directive as transposed into German law, which is immune to opt-outs from rightsholders.[24] This exception applies when the data mining is carried out for scientific research purposes, as opposed to commercial exploitation. The Hamburg court agreed with LAION, stating that even though the dataset would later be used by commercial entities, LAION itself pursued non-commercial, research-driven objectives.[25] This creates an important distinction: the non-profit entity may be shielded by the research exception, but any commercial entity using the dataset for profit cannot claim the same protection.

This raises a key point for businesses using AI models trained on datasets like LAION-5B or Common Crawl[26]. If they are commercial entities, they must be aware of any copyright reservations. The dataset itself, created under the research exception, does not include rights reservations from the original creators, such as Kneschke. This puts commercial users at risk of infringing copyright unless they

---

[22] *Kneschke v. LAION gemeinnütziger e.V.,* (case No. 310 O 227/23).

[23] LAION (Large-scale Artificial Intelligence Open Network) is a non-profit organisation dedicated to advancing open-access research in machine learning and artificial intelligence. Its LAION-5B dataset is a large-scale collection of over five billion image-text pairs, scraped from publicly accessible online sources, intended for use in training multimodal AI models.

[24] §§ 44a, 44b and 60d UrhG

[25] *Kneschke v. LAION gemeinnütziger e.V.,* (case No. 310 O 227/23) at paragraph 114

[26] Common Crawl is a non-profit organisation that creates and maintains an open repository of web data, enabling research and innovation in fields such as natural language processing and artificial intelligence. Its dataset is a large-scale collection of raw web page data, metadata, and text extractions, compiled from periodic crawls of the public internet. Widely used in AI training, this dataset often includes publicly accessible content, raising questions about compliance with copyright law.

take steps to identify and respect those rights, which may be obscured due to the research-focused creation of the dataset.

The court also touched on the issue of opt-outs under Section 44b of the UrhG, which allows rightsholders to reserve their rights in a machine-readable format (the German transposition of Article 4 of the Copyright Directive). While Kneschke had done so in the terms and conditions of the website where his image was hosted, the court found that LAION's reliance on the research exception under Section 60d (the Article 3 transposition) rendered this irrelevant.[27] However, the court did offer some non-binding remarks on the issue, suggesting that reserving rights solely through website terms may not suffice under Section 44b.[28] This raises significant concerns for companies relying on terms and conditions as their primary method of asserting copyright protections.

The ruling suggests that more robust mechanisms, such as robots.txt files or metadata, may be necessary to assert copyright effectively. For tech companies, this introduces a new layer of complexity. If terms and conditions alone are not enough, companies may need to employ more sophisticated, machine-readable methods to ensure their rights are respected. This could be particularly challenging for organisations that have not yet adapted their copyright protections to reflect the nuances of modern AI training.

Finally, the decision signals a broader trend where the correct legal target may not be scraping entities like LAION but rather commercial companies using the datasets for profit. These companies will need to navigate the intricate interplay between copyright law and AI development. The burden of ensuring compliance with Article 53 of the AI Act, which requires commercial AI developers to account for copyright reservations, will likely increase. This makes compliance with the research TDM exception a complex endeavour, particularly as datasets like LAION's lack the necessary rights reservations to comply with copyright obligations for commercial use.

The Hamburg court's decision sets an important precedent in Europe's emerging AI regulatory framework. While non-profit entities like LAION may be able to invoke the research exception, commercial entities need to tread carefully. They must ensure they can track and comply with rights reservations when using datasets derived under research exceptions. The decision raises more questions than it answers about the future of AI and copyright, but it is clear that companies must stay vigilant in an era of increasingly sophisticated data mining practices.

VIII. IMPLICATIONS FOR AI DEVELOPERS AND COPYRIGHT HOLDERS

The EU AI Act introduces significant new obligations for AI developers, particularly those working with GPAI models.[29] The requirement to comply with EU copyright law, even for models trained outside the Union, presents a major

[27] *Kneschke v. LAION gemeinnütziger e.V.,* (case No. 310 O 227/23) at paragraph 109
[28] *Kneschke v. LAION gemeinnütziger e.V.,* (case No. 310 O 227/23) at paragraph 105
[29] Artificial Intelligence Act, *supra* note 8.

challenge for companies seeking to deploy AI systems in the EU market.[30] Developers must now navigate the complexities of both copyright law and AI regulation, and they must ensure that they have the necessary permissions to use copyrighted materials in AI training.

For copyright holders, the AI Act offers new avenues to protect their works from unauthorised use in AI systems. The transparency requirements, combined with the obligation for AI providers to comply with copyright reservations, provide rightsholders with a clearer path to enforcing their rights and seeking compensation for unauthorised use.

However, the Act also raises important questions about how copyright compliance will be enforced in practice. With the rise of large, multinational AI providers, enforcement may prove challenging, particularly in cases where the training data is sourced from jurisdictions with different copyright laws.

It remains to be seen how the LAION decision will impact AI development in the EU context. Many of the massive datasets used to train our most powerful AI models such as the Common Crawl and LAION 5B have been created by non-profit research organisations. These research organisations appear to be permitted to carry out TDM on copyright works pursuant to Article 3 of the 2019 Copyright Directive, and it would appear that in certain instances, the datasets may have been lawfully created in compliance with EU law.

The problem arises when an entity then comes along which utilises those databases for commercial purposes. If an entity carries out TDM on one of these research created databases for commercial purposes, Article 4 of the 2019 Copyright Directive applies, and such an entity would need to take heed of any reservations of rights against TDM by rightsholders. However, such reservations of rights would usually only be attached, for example, to the website on which the copyright content originally resided, be it through robots.txt files or website terms. These datasets likely do not contain reservations of rights, and there is an open question as to whether providers of GPAI models under the AI Act will be able to rely on these datasets to train AI models going forward, if they are not able to guarantee compliance with EU copyright law and adherence to rights reservations when it comes to TDM.

IX. A POTENTIAL INTERMEDIATE SOLUTION

In response to these concerns, Professor Alexander Peukert has suggested an "intermediate solution" to the problem of extraterritorial copyright enforcement under the AI Act. Peukert was recently appointed as chair of the working group

[30] *See, e.g.,* Paul Goldstein, Christiane Stuetzle, & Susan Bischoff, *Copyright Compliance With the EU AI Act — Extraterritorial Traps for the Unwary*, BLOOMBERG LAW: PRACTICAL GUIDANCE (June 2024), https://www.bloomberglaw.com/external/document/X5IQKDK000000/tech-telecom-professional-perspective-copyright-compliance-with-.

on the EU's GPAI Code of Practice on copyright. In a recent article,[31] Peukert acknowledged the concerns around the AI Act's potentially far-reaching effects and hints at an "intermediate solution" to address this challenge.[32]

In his paper, Peukert has extensively discussed the tension between exclusivity and inclusivity in intellectual property.[33] He suggests that global AI training, which often involves massive datasets scraped from the internet, requires a careful balance between these two principles.[34] He argues that while the EU should indeed protect its copyright holders from having their works used without permission, there must be room for AI providers outside the EU to operate without being hamstrung by conflicting legal regimes.[35]

Peukert's intermediate solution seems to be a proposal to limit the extraterritorial application of EU copyright law to cases where scraping and training activities target EU-hosted websites. This would mean that if an AI provider is scraping content hosted on servers in the EU, even if the training occurs outside of the EU, they would still need to comply with the copyright reservations expressed by EU rightsholders. This approach would reduce the burden on non-EU developers while still protecting EU-based content from unauthorised use in AI training.[36]

Such a solution would reflect a more targeted approach to extraterritoriality, focusing specifically on content with a clear connection to the EU. Peukert's proposal acknowledges the principle of territoriality that underpins international copyright law, while also recognising that the global nature of AI training requires nuanced regulation.

The EU's AI Act could mark the beginning of what some commentators have described as a regulatory "race to the top" in AI governance.[37] By enforcing its copyright standards globally—at least for models placed on the EU market—the EU is setting the stage for other jurisdictions to follow suit, potentially creating a more harmonised global approach to AI and copyright law. However, the challenge lies in balancing this ambition with the practical realities of AI development, which is inherently global and relies on the free flow of data.

Peukert's suggestion of an intermediate solution offers a way forward, but it also raises new questions. Will this approach be enough to satisfy non-EU

---

[31] Alexander Peukert, *Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copy* (July 25, 2024) (Research Paper of the Faculty of Law, Goethe University Frankfurt am Main), No. 3/2024, https://ssrn.com/abstract=4905400.

[32] *Id.* ¶ 19.

[33] *See*, *e.g., id.* ¶¶ 4, 6, 7, 8, 11. For Peukert's main salient points regarding the conceptual debate, *see id.* ¶ 22.

[34] *Id.* ¶ 2 (noting that patents are accompanied with this notion of "two-sided rights," which permits the holder to fully exclude one party, while simultaneously authorizing use for others).

[35] *Id.* ¶¶ 19, 20.

[36] *Id.*

[37] *See* Nathalie Smuha, *From a 'Race to AI' to a 'Race to AI Regulation': Regulatory Competition for Artificial Intelligence*, 13 LAW INNOVATION & TECH. 57 (2021).

developers? How will enforcement work in practice? And will other jurisdictions—such as the U.S., which has a more permissive approach to text and data mining—respond with similar measures?

CONCLUSION

The EU AI Act represents a significant step forward in the regulation of artificial intelligence, particularly in its approach to copyright law. By requiring AI developers to comply with EU copyright standards, the Act seeks to protect the rights of creators while fostering innovation in AI technology. However, the extraterritorial application of these rules raises complex legal and operational challenges for AI providers, particularly those operating outside the EU.

As the global AI landscape continues to evolve, it is likely that other jurisdictions will follow the EU's lead in establishing similar regulatory frameworks. The challenge for AI developers will be to navigate these overlapping legal regimes while continuing to innovate and deploy AI systems on a global scale. For copyright holders, the AI Act offers new opportunities to enforce their rights and protect their works in an increasingly digital world. The future of AI and copyright will undoubtedly be shaped by the ongoing interplay between innovation, regulation, and intellectual property law.