

Faculty of Law
Research Paper No. 3/2024

**Regulating IP Exclusion/Inclusion on a Global Scale:
The Example of Copyright vs. AI Training**

*Alexander Peukert**

Forthcoming in: liber amicorum Geertrui van Overwalle

Proposed citation: Peukert, Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copyright vs. AI Training, Research Paper of the Faculty of Law of Goethe University Frankfurt/M. No. 3/2024, para.

Abstract: This article builds upon the literature on inclusion/inclusivity in IP law by applying these concepts to the example of the scraping and mining of copyright-protected content for the purpose of training an artificial intelligence (AI) system or model. Which mode of operation dominates in this technological area: exclusion, inclusion or even inclusivity? The features of AI training appear to call for universal and sustainable “inclusivity” instead of a mere voluntary “inclusion” of AI provider bots by copyright holders. As the overview on the copyright status of AI training activities in different jurisdictions and emerging laws on AI safety (such as the EU AI Act) demonstrates, the global regulatory landscape is, however, much too fragmented and dynamic to immediately jump to an inclusive global AI regime. For the time being, legally secure global AI training requires the voluntary cooperation between AI providers and copyright holders, and innovative techno-legal reasoning is needed on how to effectuate this inclusion.

* Dr. iur., Professor of Civil Law, Commercial Law and Information Law, Goethe University Frankfurt am Main, Faculty of Law.

I. Introduction

- 1 Exclusivity is the key feature of intellectual property (IP) rights in general, and patents in particular. The rightsholder has the exclusive power to either authorize or prohibit the use of the protected subject matter by third parties, thereby excluding them from the benefits of the IP.
- 2 The concept of exclusivity in IP law has been a major theme in Geertrui van Overwalle's oeuvre. In her 2012 article "Individualism, Collectivism and Openness in Patent Law", she lays out how individual licensing schemes and collaborative/collective license structures such as patent pools and clearinghouses can "moderate the effect of IP exclusivity and turn the ... ownership regime into (semi-)open infrastructures".¹ This observation points to the fact that patents are two-sided rights, which empower the patent holder, on the one hand, to exclude others and, on the other, to include others by authorizing uses.² Thus, IP rights cannot be reduced to their exclusionary effect alone. They also form the basis for very flexible and granular modes of sharing, exchange, financing and risk-spreading.³
- 3 Contractual inclusion is not a panacea though. Rightsholders are generally free to deny authorization of a requested use, and they can demand a royalty in an amount determined at will. In certain situations, the possibility of obtaining a license does not provide a sufficient solution to accommodate a legitimate interest to access and use protected IP.⁴ Even if a rightsholder dedicates her invention to the public domain, problematic exclusionary effects may still return and prevent universal and sustainable access. In particular, follow-on innovations that build on the public domain can be re-commodified.⁵

¹ Geertrui Van Overwalle, 'Individualism, Collectivism and Openness in Patent Law' in Jan Rosén (ed), *Individualism and Collectiveness in Intellectual Property Law* (Edward Elgar 2012), 71-116.

² Geertrui Van Overwalle, 'Inventing Inclusive Patents. From Old to New Open Innovation' in Peter Drahos, Gustavo Ghidini and Hanns Ullrich (eds), *Kritika: Essays on Intellectual Property* vol. 1 (Edward Elgar 2015), 206, 250.

³ Generally Daniel B Kelly, 'The Right to Include' (2014) 63 Emory Law Journal 857, 922.

⁴ Regarding artistic sampling see German Federal Constitutional Court, Judgment of 31 May 2016, case 1 BvR

1585/13 – Metall auf Metall I <https://www.bverfg.de/e/rs20160531_1bvr158513en.html> accessed 20 June 2024, para 98.

⁵ Geertrui Van Overwalle, 'Inventing Inclusive Patents. From Old to New Open Innovation' in Peter Drahos, Gustavo Ghidini and Hanns Ullrich (eds), *Kritika: Essays on Intellectual Property*, vol. 1 (Edward Elgar 2015), 206, 232-233; Séverine Dusollier, 'The Commons as a Reverse Intellectual Property - From

- 4 These limits of voluntary inclusion have given rise to a rich and growing body of scholarship that looks for alternative ways to make IP law more “inclusive”.⁶ Geertrui van Overwalle contributed a ground-breaking article to this strand of literature by inventing an “inclusive patent”, which would empower its owner only to include others via open-source-type licenses, thereby establishing neutral and non-discriminatory, universal and perpetual access and use.⁷
- 5 It is accordingly important to distinguish between “inclusion” and “inclusivity”. While the former term describes an act of including someone or something, the latter denominates a quality of something being inclusive.⁸ Studies on different modes of inclusion also tend to be descriptive, whereas the inclusivity literature expressly adopts a more forward-looking, normative perspective that criticizes the exclusion of certain communities or individuals and instead favours a “more open”, inclusive regulatory framework for IP.⁹

Exclusivity to Inclusivity’ in Helena Howe and Jonathan Griffiths (eds), *Concepts of Property in Intellectual Property* (CUP 2013) 258, 273-4. For other shortcomings of voluntary inclusion through open source/access schemes see Séverine Dusollier, ‘Sharing access to intellectual property through private ordering’ (2007) 82 *Chicago-Kent Law Review* 1391, 1434; Séverine Dusollier, ‘Ecologies of IP Openness’ in Francois Thouvenin and others (eds), *Kreation Innovation Märkte - Creation Innovation Markets* (Springer 2024), 3-15.

⁶ Séverine Dusollier, ‘The Commons as a Reverse Intellectual Property - From Exclusivity to Inclusivity’ in Helena Howe and Jonathan Griffiths (eds), *Concepts of Property in Intellectual Property* (CUP 2013), 258, 273; Séverine Dusollier, ‘Inclusivity in intellectual property’ in Graeme Dinwoodie (ed), *Intellectual Property and General Legal Principles* (Edward Elgar 2015), 101-118; Anna Rogler, *Inklusive Immaterialgüterrechte* (Duncker & Humblot 2020); Cristiana Sappa (ed), *Research Handbook on Intellectual Property Rights and Inclusivity* (Edward Elgar 2024).

⁷ Geertrui Van Overwalle, ‘Inventing Inclusive Patents. From Old to New Open Innovation’ in Peter Drahos, Gustavo Ghidini and Hanns Ullrich (eds), *Kritika: Essays on Intellectual Property*, vol. 1 (Edward Elgar 2015), 206, 250. This proposal has been taken up and developed further by, e.g., Karen Walsh and others, ‘Intellectual property rights and access in crisis’ (2021) 52 *International Review of Intellectual Property and Competition Law* 379-416; Michael A Kock, ‘Open intellectual property models for plant innovations in the context of new breeding technologies’ (2021) 11(6) *Agronomy* 1218; Jaako Siitaloppi and Rosa Maria Ballardini, ‘Promoting systemic collaboration for sustainable innovation through intellectual property rights’ (2023) 11(1) *Journal of Co-operative Organization and Management* 100200.

⁸ Cf <<https://www.oed.com/search/dictionary/?scope=Entries&q=inclusion>>; <<https://www.merriam-webster.com/dictionary/inclusivity>> accessed 20 June 2024.

⁹ Geertrui Van Overwalle, ‘Inventing Inclusive Patents. From Old to New Open Innovation’ in Peter Drahos, Gustavo Ghidini and Hanns Ullrich (eds), *Kritika: Essays on Intellectual Property*, vol. 1 (Edward Elgar 2015) 206, 225-226 (“more open” is good for innovation”); Séverine Dusollier, ‘The Commons as a Reverse Intellectual Property - From Exclusivity to Inclusivity’ in Helena Howe and Jonathan Griffiths (eds), *Concepts of Property in Intellectual Property* (CUP 2013), 258, 275 (“counter the empty promise of the public domain”); Séverine Dusollier, ‘Inclusivity in intellectual property’ in Graeme Dinwoodie (ed), *Intellectual Property and General Legal Principles* (Edward Elgar 2015), 101, 118 (replace the foundational private/public dichotomy “with a more complex and fertile mix of rights and privileges”); Cristiana Sappa, ‘Introduction: Intellectual Property Rights and Inclusivity’, in Cristiana Sappa (ed), *Research Handbook on*

6 This article builds upon the literature on inclusion/inclusivity in IP law by applying these concepts to the example of the scraping and mining of copyright-protected content for the purpose of training an artificial intelligence (AI) system or model.¹⁰ Which mode of operation dominates in this technological area: exclusion, inclusion or even inclusivity? The example of data mining for AI training purposes is interesting because it tests the inclusion/inclusivity framework in an extreme context. First, general-purpose AI models (GPAIMs) such as Open AI's Generative Pre-Trained Transformers are typically developed and trained on "vast amounts of text, images, videos, and other data".¹¹ "The larger a data set, the better even subtle relations in the data can be discovered" and represented in the model.¹² GPAIM providers therefore typically employ web robots to collect hundreds of millions or even billions of texts, pictures and other data by automated website scraping.¹³ AI therefore represents a highly complex technology in the sense that its development regularly involves billions of potentially copyright-protected data, in contrast to a "discrete" technology like an active pharmaceutical ingredient where separate technological opportunities or even final products are subject to only one or few IP rights.¹⁴ Second, website scraping and consecutive AI training typically take place transnationally across many IP jurisdictions. Following the logic that much data is good, but more data is better, global data collections are preferable to data collections limited to a specific location or language. The territorial and cultural diversity of AI training

Intellectual Property Rights and Inclusivity (Edward Elgar 2024), 1, 8.

¹⁰ For a definition of these terms see Regulation (EU) 2024/... of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (Artificial Intelligence Act, in the following: EU AI Act), art 3(1), (29) and (63).

¹¹ EU AI Act, recital 105 2nd sentence.

¹² European Commission, 'Artificial Intelligence for Europe' (Communication) COM (2018) 237 final, 15.

¹³ Cf *HiQ Labs, Inc v LinkedIn Corp* 31 F.4th 1180, 1186-1187 (9th Cir 2022); Andres Guadamuz, 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs' [2024] GRUR International 111, 115; Roman Konertz and Raoul Schönhof, 'Vervielfältigungen und die Text- und Data-Mining-Schranke beim Training von (generativer) Künstlicher Intelligenz: Über die Codierung von Daten und Informationen in Künstlichen Neuronalen Netzen' [2024] Wettbewerb in Recht und Praxis 289, 290 (GPT-3 developed on the basis of 45 terabyte text, 159 gigabyte source code; Stable Diffusion on the basis of 160 million pictures).

¹⁴ Such a discrete technology is discussed in Geertrui Van Overwalle, 'Individualism, Collectivism and Openness in Patent Law', in Jan Rosén (ed), *Individualism and Collectiveness in Intellectual Property Law* (Edward Elgar 2012) 71. On the difference between discrete and complex technologies see Alexander Peukert, 'Virtual Patent Networks and Their Network Effects' in Christine Godt and Matthias Lamping (eds), *A Critical Mind - Hanns Ullrich's Footprint in Internal Market Law, Antitrust and Intellectual Property* (Springer 2023), 303, 318-320 with further references.

sources also has a normative aspect. The World Economic Forum has emphasized that “active inclusion of developing nations and diverse voices in generative AI development and governance is critical to ensure global inclusion in a future influenced by generative AI”.¹⁵

- 7 These features of AI training appear to call for universal and sustainable “inclusivity” instead of a mere voluntary “inclusion” of AI provider bots by copyright holders. As the following overview on the copyright status of AI training activities in different jurisdictions (infra, II) and emerging laws on AI safety (infra, III) demonstrates, the global regulatory landscape is, however, much too fragmented and dynamic to immediately jump to an inclusive global AI regime. For the time being, legally secure global AI training requires the voluntary cooperation between AI providers and copyright holders, and innovative techno-legal reasoning is needed on how to effectuate this inclusion (infra, IV).

II. The Uncertain Copyright Status of AI Training

- 8 The question of whether and under what conditions copyrighted material may be reproduced for the training of commercial AI is currently a hot topic around the world and provides an example of the recurring conflict between IP exclusivity and access interests. Broadly speaking, there are three solutions to solve the copyright vs. AI training conflict. At one extreme, scraping and mining of copyrighted content for AI training purposes could be considered an infringement if done without prior specific authorization. At the other extreme, AI training could be deemed permissible without prior authorization and payment, even if done for commercial purposes. Among the regulatory options in between these extremes, copyright law could permit reproductions for AI training under certain conditions, in particular only for non-commercial uses or on the condition of lawful access to the source or on the condition that the rightsholder has not vetoed the scraping of her content in a machine-readable manner. A brief comparative overview demonstrates that all of these solutions find support in different copyright jurisdictions.

¹⁵ World Economic Forum, ‘Generative AI Governance: Shaping a Collective Global Future’ (World Economic Forum 2024), 11. See also Maximiliano Marzetti, ‘Towards a more inclusive global public domain’ in Cristiana Sappa (ed), *Research Handbook on Intellectual Property Rights and Inclusivity* (Edward Elgar 2024), 212-228.

9 Unauthorized reproductions for AI training purposes are likely to be found infringing under laws that do not provide a specific rule for this type of use, a text and data mining (TDM) exception, or a general “fair use” clause.¹⁶ A draft bill for an amendment of the Polish Copyright Act even explicitly states that the Polish TDM exception does not apply to “the creation of generative artificial intelligence models, unless otherwise stipulated by the authorized party”.¹⁷ The opposite outcome, namely the permissibility of commercial AI trainings, could be derived from the US “fair use” clause,¹⁸ and a provision in the Japanese Copyright Act legalizing commercial uses of a work “in a data analysis”, provided that this exploitation does not unreasonably prejudice the interests of the copyright owner.¹⁹ An example for an intermediate solution can be found in Art. 4 of the EU Directive 2019/790 on Copyright in the Digital Single Market (CDSMD),²⁰ which legalizes reproductions and extractions of lawfully accessible works and other subject matter for the purposes of commercial TDM, subject to the condition that the use “has not been expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online”. It is, however, argued that this provision is not applicable to AI training activity in the first place,²¹ and it is moreover unclear under which conditions “reservations of rights” are in fact “expressed pursuant to” paragraph 3 of this provision. Can an opt-out be expressed *ex post*²² and with regard to a work as such

¹⁶ For Brazil see Florence G’sell, *Regulating under Uncertainty: Governance Options for Generative AI* (4 June 2024) 372-374.

¹⁷ Paul Keller, ‘TDM: Poland challenges the rule of EU copyright law’ (*Kluwer Copyright Blog*, 20 February 2024) <<https://copyrightblog.kluweriplaw.com/2024/02/20/tdm-poland-challenges-the-rule-of-eu-copyright-law/>> accessed 8 April 2024. See also Haimo Schack, ‘Auslesen von Webseiten zu KI-Trainingszwecken als Urheberrechtsverletzung de lege lata et ferenda’ [2024] *Neue Juristische Wochenschrift* 113, 114 (Art 4 EU Directive 2019/790 does not apply to commercial AI training).

¹⁸ See 17 USC §107 and Matthew Sag, ‘Fairness and Fair Use in Generative AI’ (2024) 92 *Fordham Law Review* 1887.

¹⁹ Cf Tatsuhiro Ueno, ‘The Flexible Copyright Exception for “Non-Enjoyment” Purposes – Recent Amendment in Japan and Its Implication’ [2021] *GRUR International* 145; but see Agency for Cultural Affairs, Government of Japan, ‘AI and Copyright Concepts’ (15 March 2024) <https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/hoseido/r05_07/pdf/94024201_01.pdf>, <https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf> accessed 20 June 2024 (the proviso should apply if the AI output competes with the works used for the training and if technical opt-outs are circumvented or ignored).

²⁰ Directive (EU) 2019/790 on copyright in the digital single market (CDSMD) [2019] OJ L 130/92.

²¹ Haimo Schack, ‘Auslesen von Webseiten zu KI-Trainingszwecken als Urheberrechtsverletzung de lege lata et ferenda’ [2024] *Neue Juristische Wochenschrift* 113; contra Alexander Peukert, ‘Copyright in the Artificial Intelligence Act – A Primer’ [2024] *GRUR International* 497, 503-4 with further references.

²² Péter Mezei, ‘A saviour or a dead end? Reservation of rights in the age of generative AI’ (2024) 46(7) *European Intellectual Property Review* 461, 465 ff (ex ante source-page opt-outs vs. ex post training-data

and not only a particular digital representation/copy?²³ In which machine-readable manner must an Art. 4(3) CDSMD reservation be articulated in order to be considered “appropriate”?²⁴

- 10 What these comparative remarks show is that the only thing certain on the global AI training market is that there is no legal certainty. A global consensus on how to regulate the conflict between copyright holders and AI providers does not exist. Solutions, including more inclusive, i.e. access-beneficial approaches, will only be worked out by the courts and national/regional legislators in the years to come.

III. Emerging AI Safety Laws

- 11 If AI training were subject only to conventional copyright laws, AI providers could still operate effectively, despite the uncertainties just described. The reason is that according to the universally accepted principle of territoriality and the corresponding rule of *lex loci protectionis*, reproductions in the course of AI model training are governed solely by the copyright law of the country in which these reproductions take place.²⁵ Thus, if an AI provider trains a generative AI model exclusively on servers located in one country, the respective reproductions are governed solely by the copyright law of that country. As soon as courts in the country of training confirm that the reproductions are permissible, the respective models, which do not themselves contain copies of works used for the training,

opt-outs).

²³ Paul Keller and Zuzanna Warso, ‘Defining best practices for opting out of ML training’ (*Open Future*, 28 September 2023) <<https://open-future.eu/publication/defining-best-practices-for-opting-out-of-ml-training/>> accessed 27 March 2024, 12; compare also Thomas Margoni and Martin Kretschmer, ‘A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology’ [2022] *GRUR International* 685, 690 (‘the content is not protected in its own right, the container is’) with SACEM, ‘Pour une intelligence artificielle vertueuse, transparente et équitable, la Sacem exerce son droit d’opt-out’ <<https://societe.sacem.fr/actualites/notre-societe/pour-une-intelligence-artificielle-vertueuse-transparente-et-equitable-la-sacem-exerce-son-droit>> accessed 7 April 2024 (public announcement of opt-out for the complete SACEM repertoire).

²⁴ Cf CDSMD, recital 18 2nd sentence (‘including metadata and terms and conditions of a website or a service’); Péter Mezei, ‘A saviour or a dead end? Reservation of rights in the age of generative AI’ (2024) 46(7) *European Intellectual Property Review* 461, 465 (‘doctrinal and practical minefield’); David Bomhard and Jonas Sigmüller, ‘AI Act – das Trilogergebnis’ [2024] *Recht Digital* 45, 50 with reference to the definition of ‘machine-readable format’ in Directive (EU) 2019/1024 on open data and the re-use of public sector information (recast) [2019] OJ L172/56, art 2(13).

²⁵ Regarding search engines see German Federal Court of Justice, judgment of 29 April 2010, Case I ZR 69/08 – Vorschaubilder I.

could be released worldwide in a copyright-compliant manner. Rightsholders could sue an AI provider only before the courts in the country where the training took place – with little prospect of success.²⁶ AI providers, in contrast, could select an AI-friendly copyright jurisdiction for their development activities and thereby trigger a regulatory race to the bottom.²⁷

- 12 This scenario, however, already appears less likely in light of emerging AI safety laws, which will further complicate the copyright/AI issue. Numerous jurisdictions have adopted or are currently contemplating laws or regulations addressing the risks of AI for various individual and public interests.²⁸ China and the EU are at the forefront of these regulatory efforts. In August 2023, “Interim Measures on the Management of Generative Artificial Intelligence Services” entered into force in China,²⁹ and on 13 June 2024, the Presidents of the European Parliament and of the EU Council signed the EU Artificial Intelligence Act, which is expected to enter into force in the second half of 2024.³⁰
- 13 The purpose these AI safety laws is to promote the uptake of generative and other AI while at the same time ensuring that only legally compliant and safe/trustworthy/healthy models and systems are released to the public.³¹ The regulatory approach of these laws differs fundamentally from that of conventional copyright law. Like other private rights, copyrights empower each rightsholder to decide autonomously in individual cases who is permitted which use and under what conditions. In contrast, the Chinese AI Interim Measures and the EU AI Act do not follow this property logic, but rather the public law

²⁶ See generally Alexander Peukert, *Urheberrecht und verwandte Schutzrechte* (CH Beck 2023), § 49 paras 15 ff; Franz Hofmann, ‘Zehn Thesen zu Künstlicher Intelligenz (KI) und Urheberrecht’ [2024] *Wettbewerb in Recht und Praxis* 11, 14; contra Haimo Schack, ‘Auslesen von Webseiten zu KI-Trainingszwecken als Urheberrechtsverletzung de lege lata et ferenda’ [2024] *Neue Juristische Wochenschrift* 113, 114, 115 (law of the country where the AI developer is domiciled).

²⁷ Claims that AI output infringes copyright would, however, still be subject to the laws of each country in which the model is distributed.

²⁸ For an overview of approaches see World Economic Forum, ‘Generative AI Governance: Shaping a Collective Global Future’ (World Economic Forum 2024).

²⁹ 生成式人工智能管理暂行办法 <http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm> accessed 20 June 2024 (in the following: Chinese AI Interim Measures); Sara Migliorini, ‘China’s Interim Measures on generative AI: Origin, content and significance’ (2024) 53 *Computer Law & Security Review* 105985 <<https://doi.org/10.1016/j.clsr.2024.105985>> accessed 21 June 2024.

³⁰ EU AI Act, art 113.

³¹ Chinese AI Interim Measures, art 1; EU AI Act, art 1(1).

logic of legal obligations. AI safety laws are intended to protect various public interests and individual rights of AI users by imposing preventive, general and abstract obligations on actors in the AI value chain.³² They can thus be described as horizontal meta-regulation in the public interest.³³

- 14 Nonetheless, IP rights and copyrights in particular feature prominently in the AI safety laws of China and the EU. According to Art. 4(3) of the Chinese AI Interim Measures, the provision and use of generative AI shall respect IP rights. In a similarly broad manner, the EU AI Act obliges providers of GPAIMs to put in place a policy to comply with Union copyright law.³⁴ In addition, both laws establish copyright-related obligations for the AI training phase. Under Art. 7(1) and (2) of the Chinese AI Interim Measures, training data processing activities shall only use lawful sources and not infringe IP rights.³⁵ The EU AI Act establishes a more concrete obligation by stipulating that providers of GPAIMs shall “identify and comply with, including through state of the art technologies, a reservation of rights expressed pursuant to Art. 4(3) of Directive (EU) 2019/790”.³⁶ In other words, the EU AI Act obliges AI model providers to respect (and not circumvent or otherwise ignore) machine-readable TDM opt-outs, i.e. the decision of copyright holders to not have their protected subject matter used for the training of AI models.
- 15 One of the questions raised by these meta-regulations of copyright interests concerns their territorial scope of application. For example: if an AI provider based in Japan trains a model on servers in Japan using content scraped exclusively from websites hosted in Japan, will she still have to prove that the training complies with EU copyright law when she places her model on the EU market? The answer to this question is far from clear.

³² Chinese AI Interim Measures, art 4; EU AI Act, art 2(1)(a) and (e).

³³ Alexander Peukert, ‘Copyright and the meta-regulation of intermediary services and artificial intelligence’ (*Kluwer Copyright Blog*, 13 June 2024) <<https://copyrightblog.kluweriplaw.com/2024/06/13/copyright-and-the-meta-regulation-of-intermediary-services-and-artificial-intelligence/>> accessed 21 June 2024.

³⁴ EU AI Act, art 53(1)(c).

³⁵ Sara Migliorini, ‘China’s Interim Measures on generative AI: Origin, content and significance’ (2024) 53 *Computer Law & Security Review* 105985 <<https://doi.org/10.1016/j.clsr.2024.105985>> accessed 21 June 2024.

³⁶ Art 53(1)(c) EU AI Act; see Alexander Peukert, ‘Copyright in the Artificial Intelligence Act – A Primer’ [2024] *GRUR International* 497, 503-6.

16 On the one hand, the EU AI Act applies to all providers “placing on the market general-purpose AI models in the Union, irrespective of whether those providers are established or located within the Union or in a third country”.³⁷ The universal application of EU law in the case that the internal market or legitimate interests of persons established or located in the Union are affected is already known, in particular, from competition law and from the Digital Services Act.³⁸ This solution corresponds to the general purpose of the AI Act to “ensure a level playing field and an effective protection of rights and freedoms of individuals across the Union”.³⁹ It also follows logically from the fact that the AI Act does not apply during model development, as long as the model is not placed on the Union market.⁴⁰ Recital 106 EU AI Act confirms the universal scope of application of Art. 53(1)(c) EU AI Act by stating that “any” provider placing a GPAIM on the Union market should comply with the TDM opt-out obligation

“regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place. This is necessary to ensure a level playing field among providers of general-purpose AI models where no provider should be able to gain a competitive advantage in the Union market by applying lower copyright standards than those provided in the Union”.⁴¹

17 According to this maximalist interpretation, Art. 53(1)(c) EU AI Act applies extraterritorially to model trainings that take place in a third country.⁴² Meta-regulation overcomes the territorial limits of copyright law. The resulting rule is simple but far-reaching: if a GPAIM

³⁷ EU AI Act, art 2(1)(a); similarly Chinese AI Interim Measures, art 2(1) (“provide the public in the People's Republic of China” with generative AI services); Sara Migliorini, ‘China's Interim Measures on generative AI: Origin, content and significance’ (2024) 53 Computer Law & Security Review 105985 <<https://doi.org/10.1016/j.clsr.2024.105985>> accessed 21 June 2024.

³⁸ Cf Regulation (EC) No 864/2007 on the law applicable to non-contractual obligations (Rome II) [2007] OJ L199/40, art 6(1); Regulation (EU) 2022/2065 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) [2022] OJ L 277/1, art 2(1).

³⁹ Cf EU AI Act, recital 21 and art 1(1).

⁴⁰ Similarly Chinese AI Interim Measures, art 2(3).

⁴¹ EU AI Act, recital 106 3rd and 4th sentence. See also European Parliament, ‘Report on cultural diversity and the conditions for authors in the European music streaming market’ 2023/2054(INI), para 27; David Bomhard and Jonas Sigmüller, ‘AI Act – das Trilogergebnis’ [2024] Recht Digital 45, 51.

⁴² de la Durantaye (n 19) para 40; Martin Senftleben, ‘AI Act and Author Remuneration – A Model for Other Regions?’ <<http://dx.doi.org/10.2139/ssrn.4740268>> accessed 5 April 2024, 13.

has not been developed in accordance with the AI Act, it may not be placed on the EU market.⁴³

- 18 On the other hand, if the model training is not governed by Art. 4 CDSMD in the first place (because it took place in another country), there is also no “reservation of rights expressed pursuant to” Art. 4(3) CDSMD that the GPAIM provider can identify and comply with. Even if websites hosted, for instance, in Japan use access restrictions like the robots.txt protocol, these opt-outs are not expressed “pursuant to” Art. 4(3) CDSMD because this provision is – according to the universally accepted territoriality principle – inapplicable on Japanese territory. The same conclusion follows from the fact that “Union copyright law” only protects against uses, in particular acts of reproduction, that occur in the EU.⁴⁴ Thus, the obligation to “put in place a policy to comply with Union copyright law” can logically only extend to actions in the EU that are subject to Union copyright law. According to this minimalist reading of Art. 53(1)(c) EU AI Act, copyright law defines the territorial limits of the EU AI Act.
- 19 As an intermediate solution, one could also consider making the application of Art. 53(1)(c) EU AI Act dependent on whether the model provider scraped websites hosted on servers located in the EU. This approach has the benefit that it corresponds to the specific acts referred to in Art. 53(1)(c) EU AI Act, namely the automatic scraping of websites by bots and the exclusion of those bots by machine-readable rights reservations.⁴⁵ It would also reflect the fact that Art. 4(3) CDSMD grants rightsholders a de facto right to control access to protected content. If this content is hosted and access-controlled in the EU, it seems justified to oblige a GPAIM provider to identify and comply with the respective access controls and complementary EU laws, even if the consecutive training takes place in a third country. The problem with the lex scraping approach is that

⁴³ David Bomhard and Jonas Sigmüller, ‘AI Act – das Trilogergebnis’ [2024] *Recht Digital* 45, 46. It is beyond the scope of this article to consider whether this outcome is ‘consistent with the rules applicable at international level’; cf. Case C-18/18 *Eva Glawischnig-Piesczek v Facebook Ireland Limited* ECLI:EU:C:2019:821, para 51.

⁴⁴ Case C-523/10 *Wintersteiger AG v Products 4U Sondermaschinenbau GmbH* [2024] ECLI:EU:C:2012:220, para 25; see also Viktoria Kraetzig, ‘KI-Kunst als schöpferische Zerstörung’ [2024] *Neue Juristische Wochenschrift* 697, para 11.

⁴⁵ Cf. *HiQ Labs, Inc v LinkedIn Corp* 31 F.4th 1180, 1187 n.3 (9th Cir 2022).

the location of a server hosting protected content is a more or less random point of attachment. One can imagine, however, that rightsholders who are particularly concerned about their content being used for AI training purposes (think about press publishers) will ensure that it is stored only on access-controlled servers located in the EU. AI model providers would have to comply with these machine-readable rules or risk that the European Commission restricts the availability of the model in the EU.⁴⁶

- 20 In summary, the AI safety laws of the EU (and China) arguably extend the EU's and China's copyright laws to AI training activities taking place in a foreign country. AI providers therefore need to consider the possibility that access to the EU and/or Chinese markets may be conditional on the prior training of the AI model being in compliance with EU and/or Chinese copyright laws. In other words, AI providers can no longer be certain that training activities will not expose them to further copyright liability if the training complied with the copyright laws of the country in which the training took place. Instead, they have to align their training activities with the copyright laws of all countries in which the model will be released. Thus, AI safety laws reverse the trend that would follow from international copyright law: instead of a potential regulatory race to the bottom to attract AI development activity, AI safety laws trigger a race to the top. AI providers targeting the global market need to ensure that their training activities comply with the strictest copyright rules. Otherwise, they run the risk of being excluded from high protection markets. This effect also explains how a jurisdiction like the EU, which may not be a leader on the global AI development market,⁴⁷ can still leverage its power as an important market for the provision of AI applications to enforce and even export its normative vision of safe and trustworthy AI.

⁴⁶ EU AI Act, art 93(1)(c). Similarly Chinese AI Interim Measures, arts 20, 21. Critical regarding the effects of this policy Thomas Margoni and Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' [2022] GRUR International 685, 700; Martin Senftleben, 'AI Act and Author Remuneration – A Model for Other Regions?' <<http://dx.doi.org/10.2139/ssrn.4740268>> accessed 5 April 2024, 6 ff.

⁴⁷ Cf. European Commission, 'Communication from the European Commission on boosting startups and innovation in trustworthy artificial intelligence' COM (2024) 28 final.

IV. Conclusion

- 21 In summary, the global AI training market operates on the conventional IP logic of exclusion and voluntary inclusion, whereas specific regulatory measures to foster access to AI training data on inclusive terms could not be observed. Copyright holders defend their interest in the exclusion of AI training bots on the basis of conventional copyright law and emerging AI safety laws. The global public interest to train AI models on an inclusive range of data from different territorial and cultural sources so as to reflect the diversity of the source data in the AI output has not yet found its way into the copyright or AI safety laws analysed herein. Instead, copyright holders can employ technical measures to prohibit AI bots from scraping their content, and some laws, such as EU law, protect these voluntary exclusions against circumvention. And even if a jurisdiction would permit unauthorized reproductions of copyright-protected content for AI training purposes, this limitation of exclusivity would not automatically prohibit rightsholders from implementing technical measures to nevertheless exclude AI bots. No wonder we are seeing more and more partnerships announced between AI providers and copyright holders such as news publishers, allowing AI companies to access and use high-quality content for training purposes for a fee.⁴⁸
- 22 Regarding the conceptual debate about IP inclusion/inclusivity, the example of copyright vs. AI teaches the two things. First, inherently global activities like the training and provision of AI models are subject to the traditional logic of exclusion/inclusion because this is the basic framework established by international copyright treaties such as the Berne Convention and the WIPO Copyright Treaty. In order to make this de facto global legal system more inclusive, contracting parties to the international copyright acquis would have to agree to replace the exclusion/inclusion dichotomy “with a more complex and fertile mix of rights and privileges”.⁴⁹ In the case of copyright vs. AI, such a consensus does not exist. The more global the use, the harder it is to implement universal IP

⁴⁸ See Florence G’sell, *Regulating under Uncertainty: Governance Options for Generative AI* (4 June 2024) 105 with further references.

⁴⁹ Séverine Dusollier, ‘Inclusivity in intellectual property’ in Graeme Dinwoodie (ed), *Intellectual Property and General Legal Principles* (Edward Elgar 2015) 101, 118.

inclusivity through codification because one state can veto any respective change. Thus, the idea to make IP more inclusive through legislative change faces a serious systemic problem in a global network society. More research is needed to address this issue.⁵⁰

- 23 The second, related conclusion is that efforts to improve the inclusiveness of AI training should instead focus on appropriate means to effectively express and operationalize the will of copyright holders to allow, disallow or allow only under certain conditions the training of their content. To achieve these goals in a digital world where code is law and AI training authorizations have to be fully automated,⁵¹ conventional contract law doctrines need to be revised and possibly refined. For example, bot requests and replies expressed in robot.txt and other machine-readable files have to be translated into legally binding declarations of intent to enter into a contract.⁵² In this context, the principle that exclusive rights should come with a flexible menu of options to execute the right proves true once again. The more options the rightsholder has, the more likely it is that win-win cooperation and inclusion occur.⁵³ And the interpretation of these transactions in view of customary online practices⁵⁴ can consider legitimate interests of users to be included in

⁵⁰ The literature on IP inclusivity has not discussed cross-border uses and conflict of laws questions in detail; see also, regarding the examination of the inclusive patent, Geertrui Van Overwalle, 'Inventing Inclusive Patents. From Old to New Open Innovation' in Peter Drahos, Gustavo Ghidini and Hanns Ullrich (eds), *Kritika: Essays on Intellectual Property*, vol. 1 (Edward Elgar 2015) 206, 261-2 ("Further research should look into the possible details of arbitration and mediation within the WIPO context.").

⁵¹ Cf Alexander Peukert, 'Copyright and the Two Cultures of Online Communication' in Paul L.C. Torremans (ed), *Intellectual Property Law and Human Rights* (4th edn, Kluwer Law International 2020) 387, 401.

⁵² Regarding TDM opt-outs see supra n 22-24. Regarding search engine and news aggregator bots see Alexander Peukert, 'Schutzbereich und Fungibilität des Presseleistungsschutzrechts' [2023] *Zeitschrift für Urheber- und Medienrecht* 233-247.

⁵³ Geertrui Van Overwalle, 'Inventing Inclusive Patents. From Old to New Open Innovation' in Peter Drahos, Gustavo Ghidini and Hanns Ullrich (eds), *Kritika: Essays on Intellectual Property*, vol. 1 (Edward Elgar 2015) 206, 249-50 (inclusive patent "an additional, complementary type of patent", which expands the options of an inventor); Daniel B Kelly, 'The Right to Include' (2014) 63 *Emory Law Journal* 857, 922; Alexander Peukert, 'Immaterialgüterrecht, Privatautonomie und Innovation' in Stefan Grundmann and Florian Möslin (eds), *Vertragsrecht und Innovation* (Mohr Siebeck 2020) 69-98.

⁵⁴ Cf s 157 German Civil Code, English translation available at <https://www.gesetze-im-internet.de/englisch_bgb/> accessed 21 June 2024.

the cross-border use of the resource.⁵⁵ Accordingly, this article reconfirms the great value of contracts as a flexible “golden oldie” of resource allocation and inclusion.⁵⁶

⁵⁵ Alexander Peukert, ‘Copyright and the Two Cultures of Online Communication’ in Paul L.C. Torremans (ed), *Intellectual Property Law and Human Rights* (4th edn, Kluwer Law International 2020) 387, 401; German Federal Court of Justice, judgment of 29 April 2010, Case I ZR 69/08, paras 28 ff – Vorschaubilder I.

⁵⁶ Robert Merges, ‘Compulsory Licensing vs. the Three “Golden Oldies”. Property Rights, Contracts, and Markets’ [2004] Policy Analysis No 508.