IPD Working Paper: How to Update EU and US Copyright Regimes in the Age of AI

Authors: Roberta Carlini, Natalia Menéndez, Anya Schiffrin*

IPD Working Paper: How to Update EU and US Copyright Regimes in the Age of AI

Authors: Roberta Carlini, Natalia Menéndez, Anya Schiffrin*

Abstract: We expect the European Union (EU) and United States (US) to travel along very different paths particularly regarding the use of news and other copyrighted works in AI training and development. The two regions have different legal frameworks and traditions, with the US relying on an expansive definition of "fair use" that is not recognized in the rest of the world, while Europe offers data privacy protections that are not present in the US. Building on data provided by the European Union's Media Pluralism Monitor, this paper discusses different ways of valuing news content, the state of current copyright negotiations between news publishers and prominent AI firms, and the consequences for media sustainability, media diversity and pluralism. We note too that there are multiple proposals for statutory levies on the AI firms that would be distributed to authors. Which, if any, will be adopted is not clear.

Key words: AI, intellectual property, fair use, remuneration, EU, media pluralism

1. Introduction

When it comes to the unauthorized use of news content and other intellectual properties in the training of large language models (LLMs) by developers of artificial intelligence (AI), we expect the European Union (EU) and United States (US) to travel along very different regulatory paths. The two regions have very different legal frameworks and traditions related to copyright protections. For one, the broad definition of "fair use" that prevails in the US is not recognized in the rest of the world. Further, Europe has adopted data privacy protections for data that the US does not have. Finally, and perhaps most importantly, the most powerful AI companies are headquartered in the US, where they benefit from immense market share and outsized economic power, taking content and refusing to pay for it with near-total legal impunity.

At the time of writing, there are some 74 copyright-related lawsuits worldwide against AI companies, 63 of which are in the US. Plaintiffs include prestigious and powerful media outlets such as *The New York Times* and The Center for Investigative Reporting. The *New York Times* had filed a lawsuit against Open AI in 2023 and in December 2025 both the *New York Times* and *The Chicago Tribune* filed suits against Perplexity[1]. In September 2025, Anthropic agreed to pay out a record 1.5 billion USD after being sued for illegal downloading of hundreds of thousands of books into its LLM.[2]

In Europe, the number of court cases litigating the use of copyrighted content for training AI systems is expected to increase in the following months and in the US the judgements have mostly favored the AI firms. As well, publishers have signed deals licensing their content to AI companies for training purposes.

This paper compares the legal regimes in the United States and the European Union regarding fair use, copyright of intellectual property and how they are challenged by the development of generative AI tools. Focusing on the use of copyrighted content to train large language models in the news media sector, we analyse the state of economic relationships between news publishers and AI companies. The paper considers how existing legal regimes will affect negotiations in Europe with the development of large language models. We discuss their impact on the media business model, taking into account the different ways of calculating the market value of news. Building on the data provided by the European Union's Media Pluralism Monitor, we consider the potential consequences for media diversity and pluralism and conclude by presenting policy options.

2. Legal frameworks in the United States (US)

The legal framework in the United States is dominated by an expansive view of "fair use," a legal doctrine that allows for the limited use of copyrighted materials without the copyright owner's permission. A number of recent legal cases have focused on whether AI developers, which train their models with content scraped from the open web without the publisher's permission, fall under the "fair use" protection. In reality, the large AI developers in the US have already profited enormously from the use of other's work and have the funds to litigate repeatedly in defense of this use.

Hopes that AI would present a chance to restrict fair use or revisit some of the court rulings on copyright seem less likely. In the US there have been a number of setbacks for creators. In May 2025, the United States Copyright Office published the last of its three-part series on AI and intellectual property which said that the fair use doctrine protected some usage of copyrighted material by AI developers, but

1. Metz,Cade and Grynbaum, Michael M. "New York Times Sues AI Start-Up Perplexity Over Use of Copyrighted Work." The New York Times, December 5, 2025 https://www.nytimes.com/2025/12/05/technology/new-york-times-perplexity-ai-lawsuit.html

2. Milliot, Jim "Athropic agrees to pay 1.5 billion settlement" Publishers Weekly, Sept 8, 2025 https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/98544-anthropic-agrees-to-pay-1-5-billion-to-settle-copyright-lawsuit.html

not all. The report does not have any specifics or cite specific use cases, stating that fair use depends on 'what works were used, from what source, for what purpose, and with what controls on the outputs–all of which can affect the market.'[3] A day after the report was published, the Trump Administration fired the director of the Copyright Office, Shira Perlmutter, and it still remains whether, or how, the report will influence the dozens of active lawsuits.

In June 2025, two federal court decisions also ruled that using copyrighted content for model training is legal. In a class action lawsuit brought by three book authors against Anthropic on the training of its Claude model, the Judge ruled that the use of books in LLM training was fair use. However, the use of pirated books infringed the authors' copyright protections. Key to the decision was the ruling that the LLM had made a "transformative" change to the material it had ingested before referencing the work in its outputs.[4] In the 1994 decision on Campbell vs Acuff, the court found that creating something new offset the commercial use of what was created. In that case, parody was considered fair use.[5]

In the 2025 ruling on Kadrey v Meta[6] the judge dismissed a group of 13 authors who alleged that Meta's use of their work in the training of its LLM, "Llama," would 'dilute the market' for their work.[7][8] In the authors' amicus curiae, lawyers Jacqueline Charlesworth and Ruby A. Strassman argued that the use by Llama was not transformative, as it did not involve any criticism or commentary.[9] Instead, they argued that Llama 'algorithmically maps and stores authors' original expression so that it can be used to generate output.' Further, the lawyers argued that the transformative standard should be weighed against potential damage to the market, given that there is now a robust market for licensing content online. Therefore, the use of copyrighted material by Llama would undermine the value of the authors' work in that market. Ultimately, however, the judge in that case did not agree.

What is Fair Use? The 1976 Copyright Act

"Fair Use" is at the heart of US copyright law. Section 107 of the US Copyright Act of 1976[10] provides the statutory framework which must be met for the fair use standard to apply. There are four criteria that determine fair use under the US Copyright Act: (1) purpose and character of the use, (2) the nature of the work used, (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole, and (4) the effect on the potential market or value for the work.

To take these points in order: when it comes to the purpose and character of the use, 'courts are more likely to find non profit educational and non commercial uses are fair,' Regarding the nature of the copyrighted work, using factual work is more likely to be considered fair use than a 'creative or imaginative work.'[11] As for the "amount and substantiality of the portion used" court rulings vary. Even small amounts

3. United States Copyright Office, 'Copyright and Artificial Intelligence. Part 3: Generative AI Training.' (Library of Congress, May 2025) https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf, accessed 22 July 2025).

4. Blake Brittain, 'Anthropic wins key US ruling on AI training in authors' copyright lawsuit' (Reuters, 24 June 2025) <https://www.reuters.com/legal/litigation/anthropic-wins-key-ruling-ai-authors-copyright-lawsuit-2025-06-24/> accessed 22 July 2025.

5. United States Copyright Office, 'Copyright and Artificial Intelligence. Part 3: Generative AI Training.' Library of Congress, May 2025, p. 40, <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf> accessed 22 July 2025.

6. Kadrey v Meta [2025] Document 598

7. Blake Brittain, 'Meta fends off authors' US copyright lawsuit over AI," (Reuters, 25 June, 202) <https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/> accessed 22 July 2025.

8. Matt O'Brien and Barbora Ortutay, 'Judge tosses authors' AI training copyright lawsuit against Meta,' (PBS, 26 June 2025) <https://www.pbs.org/newshour/arts/judge-tosses-authors-ai-training-copyright-lawsuit-against-meta> accesssed 22 July 2025.

9. Kadrey v Meta [2025] Document 535

10. 17 U.S. Code § 107 - Limitations on exclusive rights: Fair use

11. 'U.S. Copyright Office Fair Use Index,' (US Copyright Office, February 2025) https://www.copyright.gov/fair-use/, accessed 23 July 2025.

may not be covered by fair use if the part used is "the heart" of the work. In theory, if the entity using the original work makes a "transformative" change, that is typically also considered fair use in that it is not simply "copying" the original.

AI developers, including social media platforms like Google, and other technology companies that profit from creative works, have long argued that their actions are "fair use" particularly given the US's broad understanding of the term. US courts have upheld this broad understanding in seminal cases such as *Authors Guild v Google* in 2015 and in 2016 on appeal.[12] Over the objections of the authors and publishers associations, Google was allowed to digitize books in part because the courts found that it was simply reproducing *parts* of the material, making it easy to find and create a "transformative" work.[13] Thus, helping make a book discoverable is considered fair use. Or as Sobel put it: Google Books and Google Images 'have been found non-infringing largely because they do not purport to be expressive works in themselves and do not resemble copyright's traditional subject matter.'[14]

Attorneys for authors, such as Charlesworth, argue that Section 107 of the US Copyright Act does not apply in the case of AI, because the entire way AI functions contravenes the four criteria laid out in Section 107.[15] According to Charlesworth, advocates for fair use of copyrighted materials in AI training claim that AI reproduction of copyrighted work is "transformative" and not designed to merely replicate large passages from training data. Charlesworth and other critics notes that AI use of copied material is based on using the expressive content of the original training material. Further, AI models 'generate copies and derivatives of training works.' She and others note that licensing and commercial use are inevitable and so the use of copyrighted material in LLMs will have an effect on the commercial market for such material.

Sobel notes the trade-offs in further restricting the use of copyrighted work: restrictions will presumably stifle innovation by limiting the availability of quality training data.[16] Lucchi believes fair use can be modified so as to solve this problem, primarily by creating exceptions for "text and data mining" (TDM) similar to those employed in Europe.[17] Dornis and Tobert point out that TDM exceptions present a whole other set of problems. We provide information below about recent court rulings allowing TDM and the debate over whether these exemptions cause legal uncertainty and are too expansive when they allow AI firms to avoid paying copyright fees. [18] TDM is a research method that uses automated software tools to gather information from vast volumes of digital data. TDM exceptions are core to European copyright legislation, which has very different standards than those of the US. With the spread of AI, those exceptions are being debated.

3. The European Union (EU) landscape after the AI Act

Prior to the passage of the Artificial Intelligence Act (AI Act),[19] the EU already had a copyright

12. Authors Guild v. Google, Inc., [2015] No. 13-4829

13. Authors Guild v. Google, Inc., [2015] No. 13-4829

14. Benjamin Sobel, 'Artificial Intelligence's Fair Use Crisis,' [2017] Colum. J.L. & Arts 45. 48

15. Jacqueline Charlesworth, 'Generative AI's Illusory Case for Fair Use' [2025] Vanderbilt Journal of Entertainment and Technology Law 323.

16. Sobel, n. 12.

17. Nicola Lucchi, 'ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems,' [2024] European Journal of Risk Regulation 15, 602-624.

18. Tim W. Dornis and Sebastian Stober, 'Generative AI Training and Copyright Law,' [2025] Transactions of the International Society for Music Information Retrieval, V(N).

19. Regulation (EU) 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 (Artificial Intelligence Act)

framework[20] in place that could be at least partially applied to the use of copyrighted content for the training of generative AI systems. Along the same lines, in 2024, the AI Act was passed with a provision on this topic.[21] However, these frameworks seem to be insufficient to keep up with reality, and as very recent case law on the matter at EU level demonstrates, important pieces are still missing. This is problematic because the copyright framework was not drafted with AI in mind and presents opportunistic and misleading interpretations of the text and data mining (TDM) exception.[22] Indeed, a July 2025 study for EU parliamentarians[23] calls for several new measures to address the gaps including "traditional knowledge" licenses and a compensation mechanism for creators. Studies commissioned by the EU Parliament are usually considered by the legislators and this could be the case in the forthcoming legislative process to reform the directive on Copyright and related rights in the digital single market (2019/790).

After the enactment of the AI Act, the AI Office launched the drawing-up process of the general-purpose AI Code of Practice. The goal of this document is to operationalize the transparency, copyright, safety and security requirements provided by the AI Act for GPAI providers. Nearly 1,000 stakeholders, from numerous professional organizations and academia attended the Code of Practice plenary kick-off event held by the AI Office in September 2024. The participants were qualified subject-matter experts who responded to the AI Office's July 2024 consultation for participation. Input from relevant stakeholders was gathered after each round of work and incorporated to some extent, although the process was later criticized for not including the views of stakeholders in support of Big Tech companies.[24]

The copyright section of the Code of Practice 'describes a set of Measures that Signatories commit to taking in order to comply with their obligation under Article 53(1)c) AI Act.'[25] According to this article, GPAI providers shall 'put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790.' The final version of the Code of Practice was launched on 10 July 2025 and established the following recommendations for GPAI developers: (1) draw up, keep up-to-date and implement a copyright policy, (2) reproduce and extract only lawfully accessible copyright-protected content when crawling the World Wide Web, (3) identify and comply with rights reservations when crawling the World Wide Web, (4) mitigate the risk of copyright-infringing outputs, and (5) designate a point of contact and enable the lodging of complaints.[26]

With this scattered and insufficient regulatory picture in mind, we will now take a look at the most relevant updates on the use of copyrighted content to training GPAI models at the EU level. In the absence of a clear and established framework, the role of jurisprudence and decisions from supervisory authorities will help to spot potential future directions, prevalent interpretations and, in other words, the EU stance towards the matter in hand.

Landmark Legal Cases in the EU

20. See Directive (EU) 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 and Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society

21. Article 53(1)(c) AI Act.

22. Articles 3 and 4 Directive 2019/790.

23. Nicola Lucchi, 'Generative AI and Copyright: Training, Creation, Regulation' (July 2025). Study requested by the European Parliament's Committee on Legal Affairs.

24. Paul Nemitz, 'How US Firms Are Weakening the EU AI Code of Practice,' (Tech Policy Press, Jun 30, 2025) <https://www.techpolicy.press/how-us-firms-are-weakening-the-eu-ai-code-of-practice/>

25. Third Draft of the General-Purpose AI Code of Practice COMMITMENTS BY PROVIDERS OF GENERAL-PURPOSE AI MODELS COPYRIGHT SECTION Introductory note by the Chair and Vice-Chair of the Copyright Section

26. European Commission, 'The General-Purpose AI Code of Practice,' 22 July 2025. <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>

In the case *French Competition Authority v Alphabet/Google* (in March 2024 the French Competition Authority fined Google 250 million EUR on the basis of French law),[27] one of the plaintiff's claims was related to the effectiveness of the right or possibility for creators to opt-out of having their content used in LLM training data:

> *With regard to "Bard", the artificial intelligence service launched by Google in July 2023, the Autorité found in particular that Bard had used content from press agencies and publishers to train its foundation model, without notifying either them or the Autorité. Google subsequently linked the use by its artificial intelligence service of the content concerned to the display of protected content, by failing to propose a technical solution for press agencies and publishers to opt out of the use of their content by Bard without affecting the display of content protected by related rights on other Google services, thus obstructing the ability of press agencies and publishers to negotiate remuneration.*[28]

This decision by the French Competition Authority represents an interesting articulation of the relationship between the pre-existing copyright legal framework and the development of GPAI models. An effective enforcement of the provision of Art. 53(1)(c) AI Act may empower the rightsholders in the negotiating process, in a market characterized by a steep imbalance of power. It would also harmonize the fragmented regulation in the EU countries and thus avoid hindering the innovation of AI technologies by GPAI developers.

In parallel to these regulatory and policy advances, a series of legal cases have been brought before the European courts. Legal action has begun to clarify how existing copyright and TDM exceptions could apply to generative AI training, and have raised contradictory interpretations.

The first landmark case was *Kneschke vs. LAION*[29] in 2024, where the German Regional Court of Hamburg found that the use of a plaintiff's[30] photograph as part of the training dataset of LAION's AI system was covered by the TDM exception for the purposes of scientific research (Section 60d German Copyright Act/Art. 3 DSM Directive). According to Sections 60d (1) and 44b (1) of the German Copyright Act (Article 2 (2) DSM Directive), the defendant's comparison with the image description to prepare its URL-based training data set was categorized as TDM. The fact that the data set was freely accessible to the public was also significantly considered. This ruling set a tone in legitimizing all parties seeking to incorporate AI training under the TDM exception from Article 3 of the DSM Directive.[31] Additionally, based on the three-step test (Article 5(5) InfoSoc Directive),[32] the Regional Court denied a narrow interpretation. The court states that the reproduction under examination can only be used to analyze the picture files for consistency with an existing image. However, a more stringent application of the three-step test to the actual AI training is unaffected by this court ruling. The LAION case is paradigmatic for the stretched interpretation offered to the notion of purposes of scientific research. Finally, the ruling will be advantageous to intermediary firms who gather training data, particularly under the court's expansive

27. Decision 24-D-03 of 15 March 2024 regarding compliance with the commitments in Decision 22-D-13 of 21 June 2022 of the Autorité de la concurrence regarding practices implemented by Google in the press sector

28. Autorité de la concurrence, press release 15 March 2024 Related rights: the Autorité fines Google €250 million for non-compliance with some of its commitments made in June 2022

29. Kneschke vs. LAION [2024] File no. 310 O 227/23. <https://openjur.de/u/2495651.html>

30. A professional photographer

31. '1. Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining. 2. Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining. 3. The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.'

32. 'The exceptions and limitations provided for in paragraphs 1, 2, 3 and 4 shall only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the rightholder.'

interpretation of Section 60d of the German Copyright Act (Article 3 DSM Directive).[33]

A German collective society for music, GEMA, was the first collective society in the world to bring legal action against a company that develops and trains generative AI models. In November 2024, GEMA sued OpenAI for reproducing German authors' protected song lyrics without paying the songwriters. The Munich Regional Court's 42nd Civil Chamber dismissed a supplementary claim pertaining to purported infringement of personality rights while generally upholding GEMA's claims for damages, disclosure, and injunctive relief. The court determined that a replication existed "in any form and by any means" in accordance with Article 2 of the InfoSoc Directive. The German legislator specifically listed "machine learning as a basic technology for artificial intelligence" within the scope of application of Section 44b German Copyright Act, and the court confirmed that training large language models will typically fall within the scope of application of the TDM barriers. However, the court determined that the reproduction of the contested song lyrics in the models does not qualify as TDM because text and data mining focuses on evaluating information like common terms, abstract syntactic rules, and semantic relationships, whereas memorization of the song lyrics in question goes beyond such an evaluation and is therefore not merely TDM.

The TDM exception was also invoked and accepted in *DPG Media et al v. HowardsHome*.[34] HowardsHome provided its consumers with articles from the Mediahuis Dutch newspaper *NRC,* collected through online RSS feeds or website scraping, rather than through purchases. The complainant contends that this violated their exclusive rights to public distribution and reproduction of their work. The court found that Mediahuis' rights reservation did not properly deny TDM permission to bots or webscrapers like the one employed by HowardsHome (para. 4.33). According to the Dutch court, an opt-out in accordance with Article 4 DSM Directive must be specific about the actors to whom the reservation is directed. The feasibility of this rule is problematic, however, since it could be regarded as an obligation for parties to list all conceivable web scrapers in the rights reservations on their respective websites.[35] The question of whether the TDM exceptions were initially intended to have such a broad scope that they could also be invoked for commercial use—such as the mass scraping by GPAI systems that we have witnessed in recent years—has been raised by a number of entities.[36]

Finally, the Municipal Court of Appeal of Hungary ruled on whether the scraping of the plaintiff's website by the top global search engine for the purpose of indexing pertinent content and providing snippet views fell under the TDM exception under Article 4 of the CDSM Directive.[37] In summary, the decision stated that search engine indexing and site scraping are "a form" of TDM. This confusion between scraping and TDM is not new. Similar stances are taken by Measures I.2.2, I.2.3, and I.2.4 in the third draft of the General Purpose AI Code of Practice. As a consequence to this decision, the Budapest Capital Regional Court has made a preliminary reference to the Court of Justice of the European Union (CJEU) in Case C-250/25, *Like Company v Google Ireland Limited.* Guadamuz has summarized the Budapest's Court arguments into the following questions:

33. Text and data mining for the purposes of scientific research.

34. Available in Dutch here: https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBAMS:2024:6563

35. Etienne Valk , Iris Toepoel, 'DPG Media et al vs. HowardsHome – A national ruling on DSM's press publishers' rights and TDM exceptions' (Kluwer Copyright Blog, 16 January 2025) <https://copyrightblog.kluweriplaw.com/2025/01/16/dpg-media-et-al-vs-howardshome-a-national-ruling-on-dsms-press-publishers-rights-and-tdm-exceptions/>

36. 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/> P. Bernt Hugenholtz, 'The New Copyright Directive: Text and Data Mining (Articles 3 and 4)' (Kluwer Copyright Blog, 24 July 2019) <https://legalblogs.wolterskluwer.com/copyright-blog/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>

37. Paul Keller, 'Do AI models dream of dolphins in lake Balaton?' (Kluwer Copyright Blog, 28 May 2025) <https://copyrightblog.kluweriplaw.com/2025/05/28/do-ai-models-dream-of-dolphins-in-lake-balaton/>

1. Does a chatbot's output that closely resembles protected parts of press publications qualify as a "communication to the public" under EU copyright law, even if the chatbot generates text through predictive modelling?

2. Does the process of training an LLM by analysing and learning from existing texts constitute a reproduction of protected works under EU law, even if it's based on pattern recognition?

3. If LLM training does count as reproduction, can it still be lawful under the text and data mining (TDM) exception in Article 4 of the DSM Directive, provided the sources were lawfully accessible?

4. If a chatbot reproduces part or all of a press publication in response to a user prompt, does that output constitute a copyright-relevant reproduction by the chatbot provider under EU law?"[38]

To conclude, the current jurisprudence at EU level pivots between denying and accepting the application of the TDM exception to the use of copyrighted content to train GPAI models, with a pending case before the CJEU whose result will have a great impact on the future interpretations by EU courts. Further, the reform of Directive 2019/790 , importantly, affects the application of copyright rules to the use of training material for GPAI models. It's not clear whether these will help to clarify the distinction between scraping and TDM, a crucial question in the current landscape.

4. Negotiating with the giants: the impact on media pluralism

AI economy and media economy

The rapid development of the generative AI systems, together with the shortcomings of the copyright legal frameworks in protecting the use of unlicensed content to train LLMs, poses a strategic choice for content creators: suing the AI companies for the use of unlicensed content, or negotiating with them to get a fair remuneration. At the time of writing, there are approximately 74 lawsuits against AI companies for copyright infringement worldwide, the majority initiated in the US.[39] In parallel, negotiations between parties started, and at the time of writing 100 agreements involving publishers and AI companies were reported. The majority of the confirmed deals have been signed with OpenAI (38) and Perplexity (21).[40] Although the economic details of the confirmed deals are not known for the majority of cases, it is estimated that as of September 2025 they were worth at least $300mns.[41] Of these deals, there are a number of high profile ones involving the media,[42] starting with the deal signed in November 2023 in Europe between Axel Springer and OpenAI, followed by *The Financial Times*, *Le Monde*, Prisa Media, News Corp, and others. In this section, we will focus on the rationale behind these negotiations for the media, considering not only the economic perspective, but also the socio-political implications for the diversity and plurality of quality media.

The impact of AI on the media and journalism is multifaceted, involving many aspects of media organisation, production and distribution. Using AI in newsrooms and the media industry opens up

38. Andres Guadamuz, 'First case on AI and copyright referred to the CJEU' (TechnoLlama, 27 May 2025) <https://www.technollama.co.uk/first-case-on-ai-and-copyright-referred-to-the-cjeu>

39. Data from <https://chatgptiseatingtheworld.com/> accessed 3 december 2025. See also, for the US: DAIL (Database of AI Litigation of the George Washington University) <https://blogs.gwu.edu/law-eti/ai-litigation-database/>

40. 'Platforms and Publishers: AI Partnership Tracker' <petebrown.quarto.pub/pnp-ai-partnerships/> accessed 3 december 2025; see also for another tracker of litigations and deals between publishers and platforms, PressGazette: and Press Gazette <https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/>

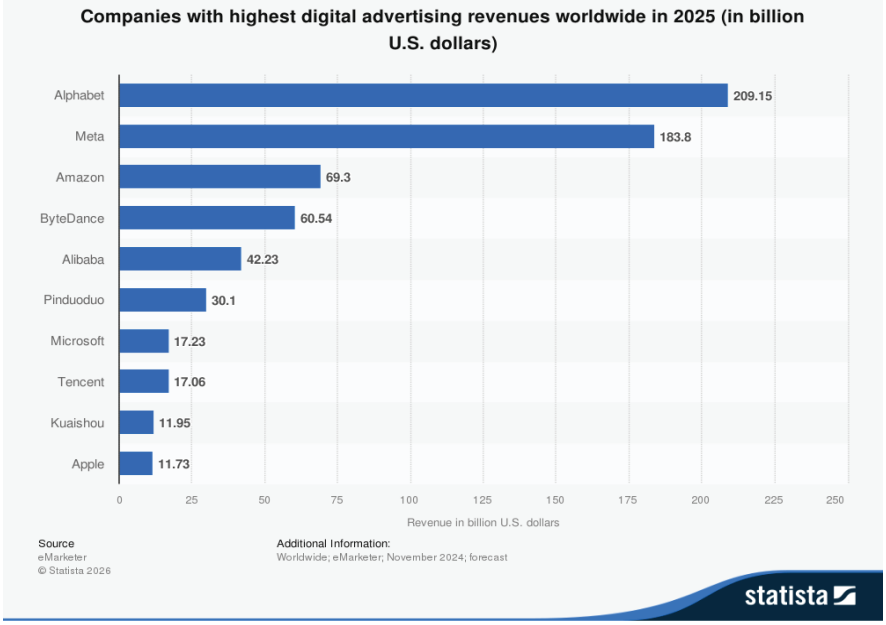
41. Peukert C., The economics of copyright and AI - Empirical evidence and optimal policy. Study for the European Parliament, Policy Department for Justice, Civil Liberties and Institutional Affairs Directorate-General for Citizens' Rights, Justice and Institutional Affairs PE 778.859 - December 2025 (p. 15)

42. WAN-IFRA, Innovation in News Media World Report 2024-2025; Pete Brown, 'Licensing deals, litigation raise raft of familiar questions in fraught world of platforms and publishers' (Columbia Journalism Review, 22 May 2024) <www.cjr.org/tow_center/licensing-deals-litigation-raise-raft-of-familiar-questions-in-fraught-world-of-platforms-and-publishers.php> accessed 17 July 2025.

potential for improving products and processes, reducing costs, engaging audiences, and making news more accessible and readable.[43] Together with these advantages, the incorporation of AI tools into news production workflows also creates risks of infrastructural dependence.[44]

Regarding the impact of AI on the media business model, there are some parallels between the current situation and the earlier rise of the platform economy and the impact on the media business mdel of search engines, social media, video-sharing platforms and other automated media aggregators.[45] Following an initial period during which traditional and digital media outlets attempted to compete with new digital intermediaries in the online advertising sector the overwhelming dominance of online platforms in the attention market prompted media outlets to adopt different strategies for monetising their content, primarily relying on paying audiences.

The media economy was still in the midst of this challenging transition, with media outlets struggling to find alternative business models, when generative AI emerged in 2022. In some ways, AI disruption resembles the 'first' digital disruption For example, fundamental elements of the platform economy, such as the massive use of copyrighted data and economies of scale, are also present in the AI economy. The tendency towards high market concentration is also similar, as seen below.



43. Tomás Dodds, Rodrigo Zamith, and Seth C. Lewis, 'The AI turn in journalism: Disruption, adaptation, and democratic futures'. (2025) Journalism, 0(0). https://doi.org/10.1177/14648849251343518 accessed 18 July 2025.

44. Efrat Nechushtai E, 'Could Digital Platforms Capture the Media through Infrastructure?' (2018) 19 Journalism 1043 <journals.sagepub.com/doi/10.1177/1464884917725163> accessed 18 July 2025; Felix M. Simon, 'Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy' (2022) 10 Digital Journalism 1832 <www.tandfonline.com/doi/full/10.1080/21670811.2022.2063150> accessed 18 July 2025. Helle Sjøvaag, 'The Business of News in the AI Economy' (2024) 45 AI Magazine 246 <onlinelibrary.wiley.com/doi/10.1002/aaai.12172> accessed 3 March 2025; Gilad Abiri, 'Generative AI as Digital Media' (Social Science Research Network, 1 March 2024) <papers.ssrn.com/abstract=4878339> accessed 18 July 2025;

45. Martin Moore and Damien Tambini (eds), Digital Dominance: The Power of Google, Amazon, Facebook, and Apple (Oxford University Press 2018;) Juan Montero and Matthias Finger, The Rise of the New Network Industries: Regulating Digital Platforms (Routledge 2021); Rasmus K. Nielsen and Sarah A. Ganter, The Power of Platforms: Shaping Media and Society (Oxford University Press 2022); Iva Nenadic, Roberta Carlini and Orlin Spassov, 'A decade of digital transformation: pluralism between the media and digital platforms' in Elda Brogi, Iva Nenadic and Pier Luigi Parcu (eds), Media Pluralism in the Digital Era (Routledge 2024). Pier Luigi Parcu, 'New Digital Threats to Media Pluralism in the Information Age' (2020) 21 Competition and Regulation in Network Industries 91 <https://journals.sagepub.com/doi/10.1177/1783591719886101> accessed 18 July 2025.; Andrea Prat and Tommaso Valletti, 'Attention Oligopoly' (2022) 14 American Economic Journal: Microeconomics 530 <www.aeaweb.org/articles?id=10.1257/mic.20200134> accessed 18 July 2025

Figure 1. Companies with highest digital advertising revenues worldwide in 2025. Bar chart showing the top 10 companies by projected global digital advertising revenue in 2025 (in billions of U.S. dollars). Alphabet leads with $209.15B, followed by Meta ($183.8B), Amazon ($69.3B), and ByteDance ($60.54B). Other companies include Alibaba, Pinduoduo, Microsoft, Tencent, Kuaishou, and Apple, with revenues ranging from $42.23B to $11.73B. The data highlights the concentration of digital advertising revenues among a few dominant global players, with U.S. and Chinese firms leading the sector. Data sourced from eMarketer, November 2024, and presented by Statista.

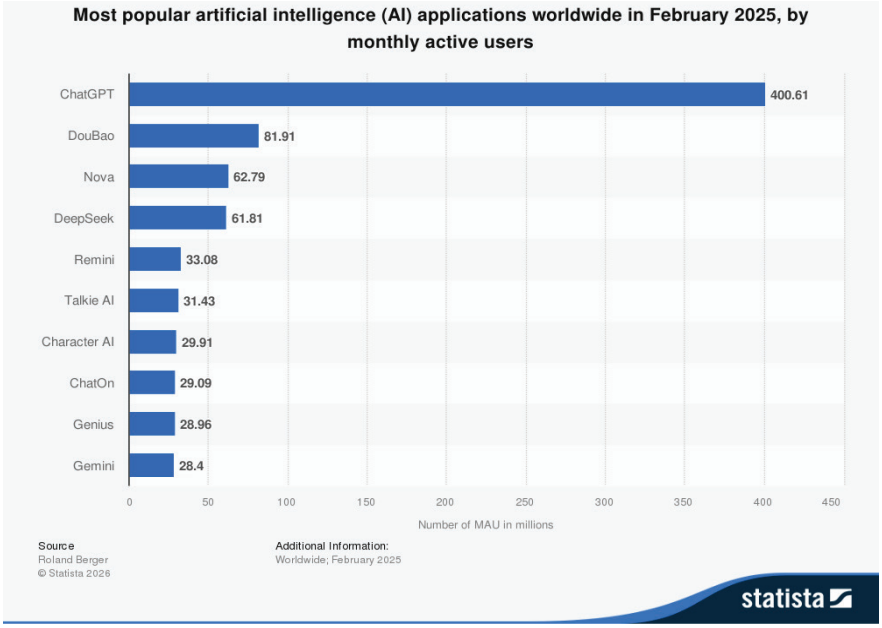


Figure 2. Most popular AI applications worldwide in February 2025, measured by monthly active users.[46] The data illustrates a steep drop-off in user base beyond the leading application, underscoring ChatGPT's dominant position in the consumer AI landscape. ChatGPT is far ahead with 400.61 million monthly active users, followed by DouBao (81.91M), Nova (62.79M), DeepSeek (61.81M), Remini (33.08M), Talkie AI (31.43M), Character AI (29.91M), ChatOn (29.09M), Genius (28.96M), and Gemini (28.4M). Data source: Roland Berger.Source: Roland Berger via Statista, February 2025.

Differences emerge when one considers the AI business model and the technology's reliance on quality media content. AI companies do not primarily generate revenue from online advertising; as their main source of revenue is the sale of their services. Even though this is expected to change,[47] currently AI companies do not compete with the media industry in the advertising market. Instead, they threaten media business models by reducing – and potentially eliminating – web traffic to media websites by providing users with automatically generated summaries of the news. There is evidence of a sharp decrease in referral traffic to media websites since the advent of AI. The more reliable and informative the automated summaries provided by AI assistants become, the less direct access to news sources will be needed or requested by news consumers. In fact, this change in news consumption habits is already underway. A study comparing search results in March 2025 with those in March 2024 found that 'the presence of an AI overview in the search results correlated with a 34.5% lower average click-through rate (CTR) for the top-ranking page,

46. Roland Berger. (March 27, 2025). Most popular artificial intelligence (AI) applications worldwide in February 2025, by monthly active users [Graph]. In Statista. Retrieved July 24, 2025, from https://www.statista.com/statistics/1609163/top-ai-applications-mau-worldwide/

47. Meghan Bobrowsky and Patrick Coffee, 'Meta Aims to Fully Automate Ad Creation Using AI' (Wall Street Journal, 2 June 2025) <www.wsj.com/tech/ai/meta-aims-to-fully-automate-ad-creation-using-ai-7d82e249>

compared to similar informational keywords without an AI overview'.[48] This trend, again is eroding media revenues by reducing the attractiveness for advertisers and revenues from subscriptions/sales.

While these automated outputs could pose an existential threat to the survival of the media, AI models rely heavily on published journalism (and, more broadly, on human-written text) to train and improve the quality of their models. Wukoson and Fortuna found that 'key LLM training datasets are disproportionately composed of high-quality content owned by commercial publishers of news and media websites.'[49]

Without quality journalism and information being fed into the models' training data, the risk of errors in the AI outputs increase. Moreover, the lack of copyright incentives may hinder the flow of the new high-quality data that is crucial for AI developers.[50] Notwithstanding broader societal considerations and regulatory obligations, this poses a risk to the AI business model itself and may explain why AI firms are willing to negotiate licensing deals with news publishers under certain circumstances. However, the use of journalism by AI firms is disrupting the media industry by reducing or nullifying the direct contact between media outlets and their audiences. Indeed, there are recent signs of growing concern about what Alex Reisner of *The Atlantic* defined as 'the end of publishing as we know it'.[51]

The value of news and AI's impact on quality journalism

The ongoing negotiations with AI companies regarding the use of media content to train LLMs, as well as the publishers' previous unsuccessful attempts to be compensated for the monetisation of their content by online platforms, have been difficult and have implications that extend beyond the economic sphere.

Assigning monetary value to news is complicated and the market has clearly failed to set a sustainable price due to information asymmetries and dominant platforms abusing their market power.[52] However, using some well-defined and widely accepted underlying principles, there are ways to calculate the dollar amounts that should be paid to news publishers. Publishers themselves sometimes use the CPM (or the cost of advertising to 1,0000 viewers) rates as a basis of comparison or how much they stand to lose by being disintermediated by large language models. This approach can be criticized as valuing the input into the model by the costs imposed on legacy media, rather than how important the news are to the model.

Haaris Mateen has attempted to estimate the value of news to the AI companies by looking at

48. Ryan Law and Xibeijia Guan, 'AI Overviews Reduce Clicks by 34.5%' (SEO Blog by Ahrefs, 17 April 2025) <ahrefs.com/blog/ai-overviews-reduce-clicks/> accessed 18 July 2025; see also: Charlotte Tobitt, 'Google AI Overviews Leads to Dramatic Reduction in Clickthroughs for Mail Online' (Press Gazette, 13 May 2025) <pressgazette.co.uk/publishers/digital-journalism/google-ai-overviews-leads-to-dramatic-reduction-in-click-throughs-for-mail-online/> accessed 16 May 2025; and Isabella Simonetti and Katherine Blunt, News Sites Are Getting Crushed by Google's New AI Tools (Wall Street Journal, 10 June 2025). "AI is killing the web. Can anything save it?" (The Economist, 14 July 2025)

49. George Wukoson and Joey Fortuna, 'The Predominant Use of High-Authority Commercial Web Publisher Content to Train Leading LLMs' (2024) <www.ssrn.com/abstract=5009668> accessed 18 July 2025

50. On the importance of focusing on data flow for value, see Peukert (2025), p. 36,

51. Alex Reisner, 'The End of Publishing as We Know It' (The Altantic, 25 June 2025) <www.theatlantic.com/technology/archive/2025/06/generative-ai-pirated-articles-books/683009/> accessed 18 July 2025

52. DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS COMPETITION COMMITTEE, 'Competition issues concerning news media and digital platforms' (OECD, 3 December 2021) <https://one.oecd.org/document/DAF/COMP(2021)16/en/pdf>; Patrick Holder et al., 'Paying for News: What Google and Meta Owe Us Publishers' [2024] SSRN Electronic Journal <https://www.ssrn.com/abstract=4704237> accessed 18 July 2025; Alexis Johann, Mia Drazilova, Sarah Treweller and Julian Möhlen, 'The value of journalistic content for the Google search engine in Switzerland' (FehrAdvice & Partners AG, March 2023) <https://fehradvice.com/insights/studien/value-of-news-study/>; Haaris Maateen and Anya Schiffrin, 'How to Calculate What News is Worth to AI' in Terry Flew, Agata Stepnik and Timothy Koskie, Palgrave (eds), Valuing News: Aligning Individual, Institutional, and Societal Perspectives (forthcoming).

share price movements of tech companies with stakes in generative AI as an indication of how the market values large language models.[53] But even if this estimate gives a broad sweep of these companies' potential revenues and profit, it doesn't provide a formula to establish the value of news articles or other words used as inputs.

Another approach is to calculate the revenues from online advertising received by a platform, e.g. Google, in a particular country, and attribute a share of it to the contribution of the media content to the overall attractiveness and audience of Google.[54] For example, during the negotiations around the Online News Act (C-18 bill) in Canada in 2023, some looked at the revenue Google generated in Canada and settled on 4% of that being given to publishers.[5556] A similar study was done by academic Jean-Hugues Roy who analyzed more than two years of Facebook data[57] During the time period analyzed, 8.9% of content posted was from news pages, suggesting that Meta owed a substantial sum to Canadian publishers. Relatedly, methodologies to determine the value of the news have been developed by Holder, Maateen, and Schiffrin [58] based on the influential study by Fehr AG for Swiss publishers.[59]

From a regulatory perspective, two Italian cases are worth mentioning. According to Italian law,[60] if the parties involved in a given licensing contract do not reach an agreement on the remuneration of copyrighted content, they can contact the Italian Authority for Communications (AGCOM), the national media and communications authority, to serve as an intermediary. In July 2024, AGCOM set the amount of fair compensation due from Microsoft to the publisher GEDI for the use of its journalistic publications in the Bing search engine.[61] In July 2025, AGCOM issued a similar decision, on the amount due by Meta platforms to GEDI, for the year 2022. In both cases, the basis for the calculation is the platforms' advertising revenues derived from the online use of the publisher's journalistic publications, net of the publisher's revenues attributable to the redirection traffic generated on its website by the journalistic publications used online by the provider.[62]

In the context of AI, it's important to note that there are different pricing structures for Retrieval Augmented Generation—which adds an external knowledge base to access up-to-date information—compared to model training and fine-tuning—which relies on a specific dataset to complete particular tasks and maintain output quality.

Media pluralism

53. Haaris Maateen and Anya Schiffrin, 'How to Calculate What News is Worth to AI' in Terry Flew, Agata Stepnik and Timothy Koskie, Palgrave (eds), Valuing News: Aligning Individual, Institutional, and Societal Perspectives (forthcoming in 2026)

54. Alexis Johann, Mia Drazilova, Sarah Treweller and Julian Möhlen, 'The value of journalistic content for the Google search engine in Switzerland' (FehrAdvice & Partners AG, March 2023) <https://fehradvice.com/insights/studien/value-of-news-study/>

55. Angwin, Julia "News Publishers are Fighting Big Tech Over Peanuts. They Could be Owed Billions" The New York Times, Dec. 8, 2023 https://www.nytimes.com/2023/12/08/opinion/google-meta-canada-press-blockade.html

56. Sam Buckingham-Jones, 'How much is news content worth to Google? Swiss researchers found out' (Financial Review, 4 September 2023) <https://www.afr.com/companies/media-and-marketing/how-much-is-news-content-worth-to-google-swiss-researchers-found-out-20230823-p5dyqk>

57. Roy, Jean-Hugues Facebook profits from Canadian media content, but gives little in return" (https://theconversation.com/facebook-profits-from-canadian-media-content-but-gives-little-in-return-146385), The Conversation, published Oct 1, 2020.

58. Patrick Holder et al., 'Paying for News: What Google and Meta Owe Us Publishers' [2024] SSRN Electronic Journal <https://www.ssrn.com/abstract=4704237> accessed 18 July 2025;

59. Alexis Johann, Mia Drazilova, Sarah Treweller and Julian Möhlen, 'The value of journalistic content for the Google search engine in Switzerland' (FehrAdvice & Partners AG, March 2023) <https://fehradvice.com/insights/studien/value-of-news-study/>

60. Legislative decree no. 177/2021, implementing EU directive (UE)2019/790;

61. Resolution 278/24/CONS

62. Art. 4(1) Annex A Resolution 3/23/CONS. On this basis a rate up to 70% applies, based on criteria set by Art. 4(2). Andrea Biondi, 'Agcom: Meta dovrà versare nove milioni di euro a Gedi' (Il Sole 24 Ore, 11 July 2025) <Agcom: Meta dovrà versare nove milioni di euro a Gedi - Il Sole 24 ORE> accessed 24 July 2025; 'Equo compenso: decisione sui diritti dovuti da Meta a GEDI' (Autorità per le Garanzie nelle Comunicazioni, 11 July 2025) <www.agcom.it/comunicazione/comunicati-stampa/comunicato-stampa-43> accessed 24 July 2025.

The results of the Media Pluralism Monitor can inform analysis of the European media environment. The MPM is a scientific tool developed by the Centre for Media Pluralism and Media Freedom, which is implemented annually in the European Union and some candidate countries. Its national results are published in country reports and analysed in a general report.[63] Adopting a holistic approach, it measures risks to media pluralism across four main areas: respect for fundamental rights, the economic dimension of market plurality, political independence, and social inclusiveness. Among the indicators used to assess the risks to market plurality, the relationships between media actors and digital gatekeepers are considered. In particular, the MPM results provide information on the state of negotiations between publishers and platforms regarding the remuneration of media content, in line with the enforcement of copyright protection, and, since the beginning of this year, the state of economic deals with AI companies. It is worth noting that the corresponding MPM variables not only ask if there are economic deals between publishers and platforms/AI companies, but also enquire into their scope and effectiveness and whether they are transparent.

According to the MPM2025 findings, in eight EU countries[64], there are no ongoing payment negotiations between publishers and platforms regarding the use of media content. In the countries where there are negotiations and agreements have been reached, these are limited. As mentioned above, in France, the competition authority stepped in the process and

> *regulatory actions intensified in 2024, notably with a €250M fine against Google for violating commitments under the 2019 law on related rights, prompting ongoing investigations into digital platforms French media launched legal offensives against X, Microsoft, and LinkedIn for unpaid content usage, while journalists' unions secured revenue-sharing agreements with Google and Meta.*[65]

As the MPM final report points out, 'some criticism emerged regarding the exclusion of the most precarious journalists from these benefits in some cases, as well as regarding the opacity of the details of the content and the amount of such agreements'.[66] In other countries, such as Denmark, publishers are offering to bargain collectively, hoping to strengthen their position in a relatively small market.[67]

In Germany, the ground-breaking deal between Axel Springer and OpenAI does not seem to have opened a path for the smaller media outlets:

> *While German market leaders signed deals with AI Companies, small and independent media outlets risk being increasingly disadvantaged when people shift from using Google Search to relying on AI models since visibility will likely favor larger media companies with licensing deals. This could lead to greater market concentration, as only a*

63. Tijana Blagojev et al, 'Monitoring media pluralism in the European Union: results of the MPM2025,' (2025) EUI, RSC, Research Project Report, Centre for Media Pluralism and Media Freedom (CMPF), <cadmus.eui.eu/entities/publication/15a6ae3c-f325-4435-a6a9-54687d595b85> accessed 18 July 2025; 'Media Pluralism Monitor' (Centre for Media Pluralism and Media Freedom, 2025) <cmpf.eui.eu/projects/media-pluralism-monitor/>, accessed 18 July 2025.

64. Namely: Croatia, Estonia, Latvia, Lithuania, Luxembourg, Poland, Romania and Slovakia.

65. Alan Ouakrat and Grégoire Bienvenu, Monitoring media pluralism in the European Union : results of the MPM2025. Country report : France. EUI, RSC, Research Project Report, Centre for Media Pluralism and Media Freedom (CMPF), (2025) 23 <hdl.handle.net/1814/92891> accessed 18 July 2025.

66. Blagojev n. 83, 59.

67. Sandra Simonsen, Monitoring media pluralism in the European Union : results of the MPM2025. Country report : Denmark. EUI, RSC, Research Project Report, Centre for Media Pluralism and Media Freedom (CMPF) (2025) <hdl.handle.net/1814/92888> accessed 18 July 2025. On 17 July 2025, the association of Dutch news media NDP Nieuwsmedia announced a partnership with the Netherlands Institute of Applied Scientific Research (TNO) to contribute to the development of GPT-NL, an open source large-scale Dutch AI language model trained on legally obtained data <ioplus.nl/en/posts/dutch-news-publishers-contribute-to-developing-gpt-nl>, accessed 19 July 2025.

few publishers gain prominence in AI-generated outputs.[68]

The risk that the AI deals only benefit the larger media outlets is also highlighted in France. After the deal signed by *Le Monde* with OpenAI,

> *APIG (Alliance de la presse d'information générale) and SEPM (Syndicat des éditeurs de la presse magazine), two press unions representing about 800 titles, asked 25 of the main AI actors (OpenAI, Google, Microsoft, Bytedance, etc.) to open negotiations on the use of their contents. Surprisingly and quite abruptly, OpenAI refused to negotiate and announced to respect the opt-out decided by these media.*[69]

In Spain, the content of the deal signed between PRISA and OpenAI in 2024 'is unknown, [including] the amount that the media will receive for their content, which contributes to the opacity of the digital media sector'.[70]

The report's findings, albeit limited in time and scope, highlight several themes. First, market and population size matter: in smaller countries and for less widely spoken languages, AI companies are very unlikely to pay for media content. Second, there is a risk of the exclusion of small and diverse media outlets. Although collective organisational processes are emerging, they have not yet produced substantial outcomes. Third, there is a lack of transparency regarding the terms and conditions of the deals, which is not necessarily justified by the need to protect industrial or commercial secrets. This is particularly concerning given the public interest in transparency in the information sphere. For example, the European Media Freedom Act introduces transparency obligations for media providers, requiring them to disclose their ownership structures and State advertising.[71] In parallel, a general interest in avoiding platform/private capture should be considered.

Proposal	Core Idea	Mechanism	Beneficiaries	Key Distinction
GRULAC (Latin American countries at WIPO SCCR)	Introduce *remuneration rights* for creators when works are used in AI training	Levy or statutory payment tied to AI training datasets	Authors and creative workers in Latin America	Focus on *redistributive justice* and curbing corporate power in global copyright debates
2025 Proposal to WIPO from Indonesia	Establish a *sovereign AI fund* to capture value from AI development	National wealth fund financed by levies on AI firms and infrastructure	State, with potential trickle-down to local creators and innovation ecosystems	Emphasizes *national development* and sovereignty rather than direct author payments

68. Jan C. Kalbhenn, Monitoring media pluralism in the European Union : results of the MPM2025. Country report : Germany. EUI, RSC, Research Project Report, Centre for Media Pluralism and MediaFreedom (CMPF), (2025) 21 <hdl.handle.net/1814/92892> accessed 18 July 2025.

69 Oukrat, n. 85, 23-24.

70 Jaume Suau Martìnez and others, Monitoring media pluralism in the European Union : results of the MPM2025. Country report : Spain. EUI, RSC, Research Project Report, Centre for Media Pluralism and Media Freedom (CMPF) (2025) 21 <hdl.handle.net/1814/92909> accessed 18 July 2025

71 Regulation (EU) 2024/1083 of the European Parliament and of the Council of 11 April 2024 establishing a common framework for media services in the internal market and amending Directive 2010/13/EU (European Media Freedom Act), Art. 6(1)(2) and Art. 25 (2).

Frank Pasquale & Haochen Sun	Proposes a streamlined opt- out mechanism and payments to those who do not opt out	Legal entitlement to compensation whenever creative works are ingested by AI	Individual creators, especially artists with distinctive styles	Also discusses levels of compensation and possible benchmarks
Christophe Geiger & Vincenzo Iaia	Advocates *statutory remuneration* for AI training	Mandatory payments whenever copyrighted works are used, replacing opt-out systems	Authors and rightsholders	Seeks a *simple, universal system* ensuring fair pay without complex licensing
EU Parliament 2025 Draft Report Rapporteur: Axel Voss	Calls for an *unwaivable right to equitable remuneration*	EU-level statutory exception allowing AI training, coupled with guaranteed payments	Authors and rightsholders across EU	Balances *innovation (AI training exception)* with *mandatory creator compensation*
Martin Senftleben	Argues for compensation due to *market substitution effects*	Copyright law reforms to provide remuneration when AI replaces human works	Human authors across journalism, music, visual arts	Focuses on *economic harm from substitution*, not just training datasets

5. Conclusion–in support of remuneration for creators and publishers

Before the Internet, the debate about copyright of intellectual property (IP) was about how to open up a closed system. At the time, it was vital to guarantee access to new ideas and make sure that scholars in underdeveloped countries could benefit from advances and innovations happening around the world. Today we discuss the opposite: how to ensure that an open system doesn't impoverish the scholars, artists, writers and creators who have devoted their lives to creating original ideas and content. In this world, it is clear that the concept of "fair use" in the United States is far too broad. The development of social media and AI and ability to copy, disseminate and profit from other people's intellectual property at scale has made a mockery of traditional justifications of fair use.

Strengthening the protection of news publishers can contribute to public access to diverse, quality information, even though this is not the main objective of copyright protection.[72] In parallel, other complementary policy tools are necessary to create a system of incentives that restore a level playing field in markets, support the public good of journalism, and address the societal harms and benefits associated with widespread adoption of generative AI.

Although the firms that are prevailing seem comfortable with "stealing and litigating, if necessary," it is possible that some will want to avoid further litigation by agreeing to pay for the use of IP—in this case, published news content. In the absence of government involvement, it is likely that some news publishers will be compensated through individual licensing arrangements with AI companies. In July 2025, the content delivery network Cloudflare announced it will collect payments from AI firms on behalf of

72. Michalina Kowala, 'Protection of Press Publishers in the Age of Generative AI – In Search of Legal Remedies to Adapt to the Pace of Technology' [2024] IIC - International Review of Intellectual Property and Competition Law <link.springer.com/10.1007/s40319-024-01515-y> accessed 15 October 2024

news publishers, as part of its services to its customers. This is an example of a private sector approach without government involvement.

The private sector involvement has one big advantage: the private sector, not the government, sets the magnitudes of the compensation. But it has one disadvantage: because the bargaining power is normally so much on the side of the AI companies, the compensation falls short of what it should be. The government can then intervene in three ways: help even the bargaining power, by organizing bargaining between large groups of publishers; provide an appellate mechanism when the outcome seems too unfair to news publishers, as Australia does; or provide itself the set of fees, after a process of evaluating the contributions of the news contribute but legacy outlets to the AI companies. In the end, simple rules may have to prevail, distinguishing less amongst quality than say might be desirable. This is what is effectively done through a fixed scale of fees that are pre-determined, as is the case with pharmaceutical licensing or music royalties.

Looking to the future and the creation of a fair system, we see four policy paths for the use of IP for AI. The first would be a free-for-all system where AI models can scrape any content they find online, and in which creators and publishers have no protection. This resembles the situation during the training period of the AI models. A second path would be one in which there is a strict no-use policy of intellectual property by AI developers beyond current exceptions afforded by existing copyright frameworks. The problem here is that the AI firms have not respected these restrictions and the courts in both the US and the EU seem to favor the AI firms' interpretation of what constitutes "fair use" or the "TDM" exception.

This path is one in which the fees to be paid are determined in the context of a negotiations framework. Above we note some of the proposals for remuneration of creators that have been proposed. We would simply add that the competitive environment has a direct bearing on how those negotiations would be conducted. A situation where powerful technology companies are on one side of the negotiating table and relatively powerless news publishers are on the other would not constitute a fair bargaining situation. For this reason, the Australian Competition Commission created the News Media Bargaining Code, understanding that the power imbalance between the negotiating partners has a significant impact on the perceived value of news.

For the negotiation framework to succeed, a number of other conditions and policies are needed. Competition law must be updated and effectively enforced to address abuse of market power (recent cases on online advertising market indicate that the competition tools are not toothless). For example, in the EU, transparency and data-sharing obligations of the digital platforms covered under the Digital Markets Act DMA must be enforced. Copyright regimes must be updated and there needs to be legal certainty and incentives to negotiate for the fair remuneration. Scholars such as Senftleben,[73] Geiger and Iaia[74] have proposed the introduction of statutory licenses to balance the interests of the industry, rightsholders, generative AI users, as well as the cultural and creative industries. In the 2025 study drafted for the European Parliament, Puekert presents the economic rationale for the statutory licensing model, which is demonstrated as a welfare-maximising policy.[75]

73. M. Senftleben, Generative AI and Author Remuneration. IIC 54, 1535–1560 (2023). https://doi.org/10.1007/s40319-023-01399-4

74. Christophe Geiger and Vincenzo Iaia, The Forgotten Creator: Towards a Statutory Remuneration Right for Machine Learning of Generative AI (October 6, 2023). Computer Law & Security Review, vol 52, 2024, 1-9., Available at SSRN: https://ssrn.com/abstract=4594873

75. "AI firms favour exceptions or no-royalty opt-outs that maximise their short-term profits, even if they risk underfunding future creation. Creators prefer statutory licensing with higher royalties, though their own surplus is maximised at an intermediate rate rather than at the highest possible one. From a social planner's perspective, a statutory licence with a modest positive royalty is usually optimal, as it balances representativeness and freshness against static distortions. However, when administrative overhead is high or pass-through from firms to consumers is weak, the welfare-maximising policy can shift toward very low or even zero royalties". Peukert C., The economics of copyright and AI - Empirical evidence and optimal policy. Study for the European Parliament, Policy Department for Justice, Civil Liberties and Institutional Affairs Directorate-General for Citizens' Rights, Justice and Institutional Affairs PE 778.859 - December 2025 (p. 26)

The proposals mentioned above of shifting to a default system of prior authorization coupled with a fixed payment scale seems to us to be the most likely to address the challenge of preserving journalism in the era of AI. Whether the political realities and pressure from the US AI firms against payment systems will prevail is another matter.

ACKNOWLEDGEMENTS: Kaylee Williams for research and editing, Joseph E. Stiglitz for editing and comments. Research support in the summer of July 2025 provided by Andrea Vega Yudico.

* Roberta Carlini is Part-Time Assistant Professor at the Centre for Media Pluralism and Media Freedom of the European University Institute. Natalia Menéndez is Assistant Professor at CUNEF University and Research Associate at the Centre for a Digital Society of the European University Institute. Anya Schiffrin is Senior Lecturer in Discipline and Co-Director of the Technology Policy and Innovation concentration at Columbia University's School of International and Public Affairs. This paper builds on the contribution to the public consultation submitted by the authors to the AI Office on the operationalisation of the transparency and copyright requirements in the AI Act.The views expressed in this essay are the authors' only and do not necessarily reflect the views of the corresponding affiliations. The authors wish to thank Andrea Vega Yudico for her research and edits.

Author contact information:

Roberta Carlini
Centre for Media Pluralism and Media Freedom, European University Institute
Badia Fiesolana
Via dei Roccettini 9, 50014
San Domenico di Fiesole (FI)
roberta.carlini@eui.eu

Natalia Menéndez
Centre for a Digital Society, European University Institute
Badia Fiesolana
Via dei Roccettini 9, 50014
San Domenico di Fiesole (FI)
natalia.menendez@cunef.edu natalia.menendez@eui.eu

Anya Schiffrin
International Affairs Building
420 W 118th St
New York, NY 10027
acs76@columbia.edu