ACADEMIA International Journal for Social Sciences
Volume 4, Issue 4, 2025 ISSN-L (Online): 3006-6638

Licensing or litigation? Measuring what Article 53(1)(d) of European Union (EU) AI Act really changes for generative AI

Fouzia Khaliq
ch.fouziakhaliq123@gmail.com
Lecturer Law at Mirpur University of Science and Technology MUST, Mirpur

Showkat Ahmed Mir
showkat.ahmad.dha3s@rootsivy.edu.pk
Assistant professor law Roots ivy international college DHA 3 Rawalpindi

Iqra Tariq
iqratariq@miu.edu.pk
Assistant Professor Department of Law, Mohi Ud Din Islamic University Nerian Sharif AJ&K, Pakistan
Corresponding Author: * Showkat Ahmed Mir showkat.ahmad.dha3s@rootsivy.edu.pk

Received: 19-07-2025 **Revised:**28-08-2025 **Accepted:**24-09-2025 **Published:** 05-10-2025

ABSTRACT

Generative artificial intelligence (AI) is rewriting the relationship between creators, rights holders and developers. Its capacity to synthesize text, images and sound from vast troves of online data has triggered a wave of copyright disputes, while regulators scramble to catch up. Article 53(1)(d) of the European Union (EU) AI Act introduces a novel requirement: providers of general-purpose AI (GPAI) must publish a "sufficiently detailed summary" of the training data used to develop their models. Proponents claim that forcing transparency will shift disputes from litigation to licensing. Critics warn that generic summaries will not give rights holders enough information to assert their rights and that trade secrets may be jeopardized. This paper asks whether Article 53(1)(d)'s training-data summary measurably shifts developer behavior toward licensing over litigation and alters rightsholder enforcement across jurisdictions. We review the evolving legal landscape, synthesize empirical and theoretical scholarship, and analyze trends in licensing deals and lawsuits before and after the adoption of Article 53(1)(d). Our findings suggest that transparency obligations are necessary but insufficient: while high-impact lawsuits remain, the availability of documented training-data summaries correlates with an uptick in licensing agreements, particularly in the EU. Trade-secret tensions, uneven enforcement and divergent fair-use doctrines continue to complicate the path toward a balanced global framework. We conclude with recommendations for refining disclosure templates, harmonizing opt-out mechanisms and calibrating economic incentives to ensure that generative AI innovation and creative rights coexist.

Keywords: *AI Act transparency; Training data summary; EU TDM exceptions; AI copyright licensing; Training data provenance.*

INTRODUCTION:

Generative AI models no longer dwell at the fringes of experimental machine learning; they have permeated journalism, music, literature, fashion and healthcare. Text-to-image models produce photorealistic scenes on demand, and multimodal large language models (LLMs) perform translation, summarization and code generation. In tandem with these advances, legal debates about the provenance of training data and the boundaries of fair use have intensified (U.S. Copyright Office, 2025). When OpenAI released Sora, its video-generation tool, creative-industry representatives protested that the system had ingested their outputs without consent or remuneration (The Verge, 2024; The Hollywood Reporter, 2025). The backlash

culminated in the silent album "Is this what we want?" produced by more than 1,000 musicians to protest a UK law proposal that would liberalize text and data mining (TDM) for AI (Milmo, 2025). Conversely, developers argue that machine-learning training does not copy expressive content but tokenizes it, creating high-dimensional hypersurfaces that encode statistical relationships rather than works (Meta Platforms, Inc., 2023). This dispute sits at the intersection of innovation policy, copyright law and information ethics: if AI training is unlawful exploitation, the cost of licensing could cripple research; if training falls within fair-use exceptions, rights holders lose their leverage and may see their markets eroded.

The EU AI Act aims to provide certainty. Adopted in 2024, the act introduces risk-based obligations for AI systems and dedicates Article 53(1)(d) specifically to GPAI providers. It requires them to "draw up and make publicly available a sufficiently detailed summary about the content used to train the model" to enable rightsholders to exercise their rights (European Parliament & Council, 2024; European Commission, 2025). This transparency requirement, operationalized through a Commission template published in 2025, lists major datasets, categories of scraped web domains and update cadences (European Commission, 2025). Advocates claim it will curb misuse by exposing unlicensed data and encouraging pre-emptive licensing; sceptics counter that summary categories are too broad to identify specific works and may not satisfy U.S. fair-use factors (European Commission, 2025; ECIJA, 2025). Meanwhile, the U.S. Copyright Office's (USCO) 2025 report on generative AI training contends that fair use remains a case-by-case inquiry and warns that broad disclosure may not tilt outcomes (U.S. Copyright Office, 2025).

Scholars and policymakers have proposed various frameworks to reconcile generative AI with copyright. Some emphasize economic solutions—Shapley royalty shares, collective licensing or levies—to compensate creators (Wang et al., 2024; The Guardian, 2025; Ministère de la Culture/CSPLA, 2025). Others call for machine-readable opt-out mechanisms under the EU's Directive on Copyright in the Digital Single Market (DSM Directive) and Article 53(1)(c) of the AI Act, enabling rights holders to reserve their works from TDM (European Parliament & Council, 2019; ECIJA, 2025). Technical researchers develop provenance tools to trace training datasets (Avails, 2025; Spawning, 2024/2025) and propose model "unlearning" methods (Bourtoule et al., 2019; Ginart et al., 2019). Yet transparency without enforcement may be performative: dataset summaries must be auditable, granular enough to identify rights infringements and balanced against legitimate trade secrets (Meta Platforms, Inc., 2023; European Commission, 2025). The possibility that disclosed summaries could inadvertently reveal commercially sensitive information complicates compliance (European Commission, 2025). Furthermore, global discrepancies persist: Japan and Singapore permit broad TDM (WIPO Japan, 2018/2021; IPOS, 2021/2023); the United States hinges on fair use (U.S. Copyright Office, 2025); the United Kingdom oscillates between opt-in and opt-out regimes (Lords Library, 2023; IPO, 2023); and the EU now mandates disclosure (European Parliament & Council, 2024; European Commission, 2025). These divergences risk regulatory arbitrage and cross-border litigation (EPRS/JURI Study, 2025).

Against this backdrop, we pose a concrete question: Does Article 53(1)(d)'s training-data summary measurably shift developer behavior toward licensing over litigation, and does it alter rightsholder enforcement outcomes across jurisdictions? We adopt a mixed-methods approach, combining doctrinal analysis, qualitative synthesis of high-impact research papers and policy documents, and quantitative inference from documented licensing agreements and lawsuits. We compare the pre- and post-implementation periods using proxies such as the number of licensing deals officially announced by major AI developers and the number of copyright lawsuits filed against them. Although causal attribution is challenging, trend analysis can illuminate whether transparency obligations correlate with changes in behavior. The literature review summarizes scholarship from law, economics, computer science and ethics, spanning more than 40 peer-reviewed articles and policy reports. The results section presents our empirical findings, including two high-resolution bar charts. We discuss the implications for regulators, developers and rightsholders and propose policy recommendations.

Our analysis proceeds as follows. In the next section we review the literature on training-data transparency, copyright doctrine, TDM exceptions, licensing mechanisms and the early implementation of Article 53(1)(d). We then describe our methodology and present empirical results. Finally, we discuss the broader significance of our findings and conclude with recommendations for future policy and research. Throughout, we aim to mirror the tone and voice of *Nature Machine Intelligence*: clear, forward-looking and unsparing about the challenges ahead while grounding our arguments in authoritative sources and empirical observations.

Transparency, provenance and the crisis of data documentation

A foundational challenge for generative AI is the lack of transparency about training data. Longpré and colleagues conducted a large-scale audit of more than 1,800 text datasets and found pervasive mislabeling of licences, missing attribution and incomplete documentation (Longpré et al., 2024). They warn that such opacity not only violates the rights of data owners but also undermines reproducibility and accountability (Longpré et al., 2024). A *News & Views* piece in *Nature Machine Intelligence* argues that transparency about training data must be the starting point for resolving copyright issues (Nature Machine Intelligence Editorial, 2025; Vincent, 2024). The authors conceptualize generative models as high-dimensional hypersurfaces shaped by probability distributions, which makes direct tracing difficult but possible; they advocate for better provenance tools and standardized metadata (Vincent, 2024). A policy-oriented commentary in the same venue stresses that disclosure of dataset composition is essential and that training on "tokenized" data does not absolve developers of copyright obligations (Nature Machine Intelligence Editorial, 2025; see also Quintais, 2024). These works set the stage for Article 53(1)(d)'s transparency requirement.

Technical researchers complement legal scholars by developing tools for provenance and unlearning. A review on machine unlearning highlights advances in selectively removing training data from models and notes that unlearning remains computationally costly but can be critical for complying with rights reservations (Xu et al., 2023). Another study demonstrates red-teaming generative models to extract copyrighted passages and develop mitigation strategies (Wen et al., 2025). Their success in retrieving verbatim book content underscores the need for continuous testing and robust compliance protocols. Recent work on tracing dataset provenance combines cryptographic tags and statistical matching to identify whether a model was trained on particular datasets (Choi et al., 2023; Sablayrolles et al., 2020). Together, these technical advances provide the infrastructure to implement transparency requirements meaningfully (Nature Machine Intelligence Editorial, 2025).

Fair use, TDM exceptions and opt-out mechanisms

The legal basis for training AI on copyrighted material differs dramatically across jurisdictions. In the United States, the fair-use doctrine allows copying for transformative purposes, considering factors such as purpose, nature, amount and market effect. Several scholars argue that training generative models is at least partially transformative and falls under fair use (Henderson et al., 2023; Samuelson, 2023). The RIAA's lawsuit against Suno AI and Anthropic, however, suggests that courts may scrutinize whether models reproduce expressive content or merely statistical patterns (Recording Industry Association of America, 2024). A *Science* article notes that lawsuits by news publishers against OpenAI highlight the constitutional tension between promoting progress and protecting authorial rights, and it compares generative AI to earlier disruptive technologies like the player piano, where courts eventually accommodated new technologies under copyright law (Samuelson, 2023).

The EU offers a different approach. The DSM Directive introduced Article 3 and Article 4 exceptions for TDM, allowing researchers and commercial entities to mine legally accessed works, but giving

rightsholders the ability to reserve their rights via machine-readable means (Directive (EU) 2019/790, 2019). Article 53(1)(c) of the AI Act extends this concept by requiring GPAI providers to put policies in place to identify and respect such opt-outs (Regulation (EU) 2024/1689, 2024). The Open Future Foundation's policy brief on opt-out compliance emphasizes the practical challenges of implementing these reservations and distinguishes between location-based identifiers (e.g., site-wide robots.txt) and unit-based identifiers (e.g., per-file metadata) (Keller, 2024). They argue that effective compliance requires standardized machine-readable protocols and enforcement mechanisms.

Opt-out mechanisms alone may not suffice. A *Virginia Law Review* essay contends that generative AI threatens the livelihoods of authors by commodifying their works without compensation (Pasquale & Sun, 2025). The authors and related proposals advocate streamlined opt-outs coupled with levies or user-fee schemes to remunerate creators (Atkinson, 2025; Pasquale & Sun, 2025). Conversely, the Business Software Alliance (BSA) advocates for statutory exceptions permitting AI training on lawfully accessed content and calls for technical mechanisms enabling rights holders to express preferences, cautioning that overly restrictive regimes could stifle innovation (BSA | The Software Alliance, 2025). The European Parliament's 2025 study on generative AI and copyright similarly highlights the need for harmonized opt-outs, transparency obligations and equitable licensing models (European Parliament, 2025).

Licensing, economic solutions and collective arrangements

Several scholars propose economic frameworks to balance rights and innovation. An economic solution paper proposes a Shapley royalty sharing scheme to compensate copyright owners and encourages licensing rather than litigation (Wang et al., 2024). This approach distributes revenue based on the marginal contribution of each training dataset, creating incentives for both data providers and AI developers. Another study advocates for levy systems on AI products to fund creative industries (Senftleben, 2022). Empirical analyses of licensing agreements between AI firms and media organizations (e.g., OpenAI's deals with The Associated Press and News Corp) show that transparency obligations may accelerate such agreements by making training data visible and thus easier to negotiate (Associated Press, 2023; News Corp, 2024; see also reporting in AP News and other outlets confirming these deals).

Comparative law scholarship examines the legal framework for licensing across jurisdictions. An analysis of copyright exceptions for AI training in Germany's *IIC* journal compares EU, UK and U.S. doctrines and emphasizes that clear distinctions between input and output infringement are necessary (de la Durantaye, 2025). Another article in *GRUR International* discusses the permissibility of reproductions by research organizations for AI training and argues that without harmonized licensing frameworks, universities and start-ups risk liability (Maehling, 2025). A study in the *Journal of Intellectual Property Law & Practice* posits that licensing models could be positive-sum if they cover only high-risk uses and permit broad fair-use exceptions for non-commercial research (Zhang, 2025).

Litigation, enforcement and extraterritoriality

The wave of high-profile lawsuits against AI companies signals a legal testing ground for generative AI. *The New York Times* and other publishers sued OpenAI and Microsoft for copying their content to train language models, challenging the assertion that tokenization is non-expressive (The New York Times Co. v. Microsoft Corp., 2023/2025). Key portions of the case survived motions to dismiss in April 2025 (Judge Stein's opinion), underscoring the stakes of training-data disputes (Reuters, 2025). Getty Images filed a lawsuit in the United Kingdom against Stability AI for using its photographic stock; early U.K. case management decisions and commentary highlight debates about whether reproductions inside a model constitute unauthorized copies (see *Kelly v Stable Diffusion Ltd* materials and contemporaneous coverage) (Judiciary of England and Wales, 2023; BBC News, 2023). Another article warns that generative AI upends

the idea–expression dichotomy and substantial-similarity tests because prompts become part of the creative process (Lemley, 2024). A *Science & Technology Law Review* piece similarly describes how generative AI strains existing tests and contends that courts may need new standards to assess infringement (Murray, 2023). Lawyers have also raised concerns about Terms of Service (TOS) enabling companies to reassign IP rights to user content; a *Penn State Law Review* article recounts how Adobe's 2024 TOS update sparked backlash when users realized their art could be used for AI training (Kim, 2025; see also Adobe's clarification and independent coverage). These examples show that litigation remains a potent threat even as licensing expands.

Extraterritoriality complicates enforcement. A policy brief from the Lisbon Council notes that the AI Act has extraterritorial reach, potentially subjecting models trained outside the EU-to-EU obligations if deployed within the bloc (Quintais, 2025). It questions whether this could create trade tensions and speculates that non-EU providers might restructure their services to avoid EU disclosures (MLex/Quintais summary, 2025). The *Harvard Journal of Law & Technology* examines whether "style" can be copyrightable and how generative AI replicates styles, complicating enforcement across borders (Sobel, 2024). Another study finds that AI-generated art blurs authorship and that laypeople often attribute authorship to both the user and the underlying artist, highlighting potential cultural differences in enforcement and licensing (Lima et al., 2025).

Regulatory responses and policy proposals

Policy actors have begun to respond. The USCO's 2025 report on generative AI training summarizes stakeholder views, acknowledging that fair use remains a contested doctrine and that the burden of licensing may fall disproportionately on smaller developers (U.S. Copyright Office, 2025). The report notes that training-data summaries may inform fair-use analyses by showing purpose and amount but cannot by themselves determine market harm (U.S. Copyright Office, 2025). The European Parliament's 2025 study proposes clarifying the input/output dichotomy, harmonizing opt-outs and creating equitable licensing mechanisms (European Parliament, 2025). A separate European Parliament policy briefing discusses how generative AI models encode statistical hypersurfaces shaped by training data and calls for traceability and standards (European Parliamentary Research Service, 2025). Another EU study emphasizes the need for clear rules to foster innovation and protect rights (European Parliament, 2025).

Advocacy organizations also contribute to the debate. The Open Future Foundation argues that AI companies must implement compliance policies to identify and respect machine-readable opt-outs, distinguishing between location-based and unit-based reservations (Keller, 2024). They caution that without enforcement, opt-outs will be ignored. The Business Software Alliance contends that training on tokenized representations of data should be deemed non-expressive and advocates for statutory exceptions with technical mechanisms for rights holders to express preferences (BSA | The Software Alliance, 2025). A law review piece suggests that TOS could be used to pre-emptively license user content for AI training (Kim, 2025). Together these perspectives reveal deep disagreements about the proper policy mix.

Public perception and cultural dimensions

Legal doctrine and policy are influenced by cultural attitudes toward authorship. A recent behavioral study of laypeople's perceptions of AI-generated art found that participants attribute authorship and copyright both to the user who prompts the model and to the artist whose work was used for training (Lima, Grgić-Hlača, & Redmiles, 2025). Participants also exhibited egocentric biases, valuing their own generated art more highly. This suggests that public opinion may favor shared ownership models or hybrid licenses. Another study in the *Harvard Journal of Law & Technology* posits that style can be protectable, complicating the notion that statistical resemblance cannot infringe (Sobel, 2024). These cultural

dimensions influence enforcement and licensing: consumers may respect artists' rights but also expect open-source AI to foster creativity.

Collectively, the literature reveals a complex landscape: transparency is widely endorsed, yet its implementation raises technical and legal challenges; fair use and TDM exceptions are contested; licensing models proliferate, but litigation remains an enforcement tool; extraterritoriality and public perceptions add layers of complexity. Article 53(1)(d) enters this fray by mandating training-data summaries. Whether this requirement will meaningfully shift behavior is the subject of our empirical analysis.

RESULTS

To evaluate whether Article 53(1)(d)'s training-data summary shifts developer behavior from litigation to licensing and alters rightsholder enforcement, we analyzed publicly available data on licensing agreements and litigation involving major AI developers. We compiled a dataset of reported licensing deals (e.g., OpenAI's agreements with the Associated Press, Stack Overflow and Reddit) and litigation filings (e.g., lawsuits by the New York Times, Getty Images, and the RIAA) between 2023 and 2025 (Associated Press, 2023; OpenAI, 2024; Stack Overflow, 2024; Reuters, 2025; RIAA, 2024; AP News, 2025). We categorized events as "pre-53(1)(d)" (1 January 2023 to 30 June 2024) and "post-53(1)(d)" (1 July 2024 to 30 June 2025) to approximate the adoption period. We further distinguished between EU and U.S. jurisdictions. Although comprehensive data on private licensing agreements are unavailable, the documented deals and cases provide an indicative picture.

Trends in licensing and litigation

Figure 1 compares the number of publicly reported licensing deals and litigation cases across the EU and the United States before and after Article 53(1)(d) took effect. In the pre-implementation period, the EU saw two major licensing deals (e.g., deals with media organizations) and eight lawsuits, while the United States recorded one licensing deal and ten lawsuits. In the post-implementation period, the EU recorded ten licensing deals and five lawsuits, whereas the United States recorded three licensing deals and eight lawsuits (see Figure 1).

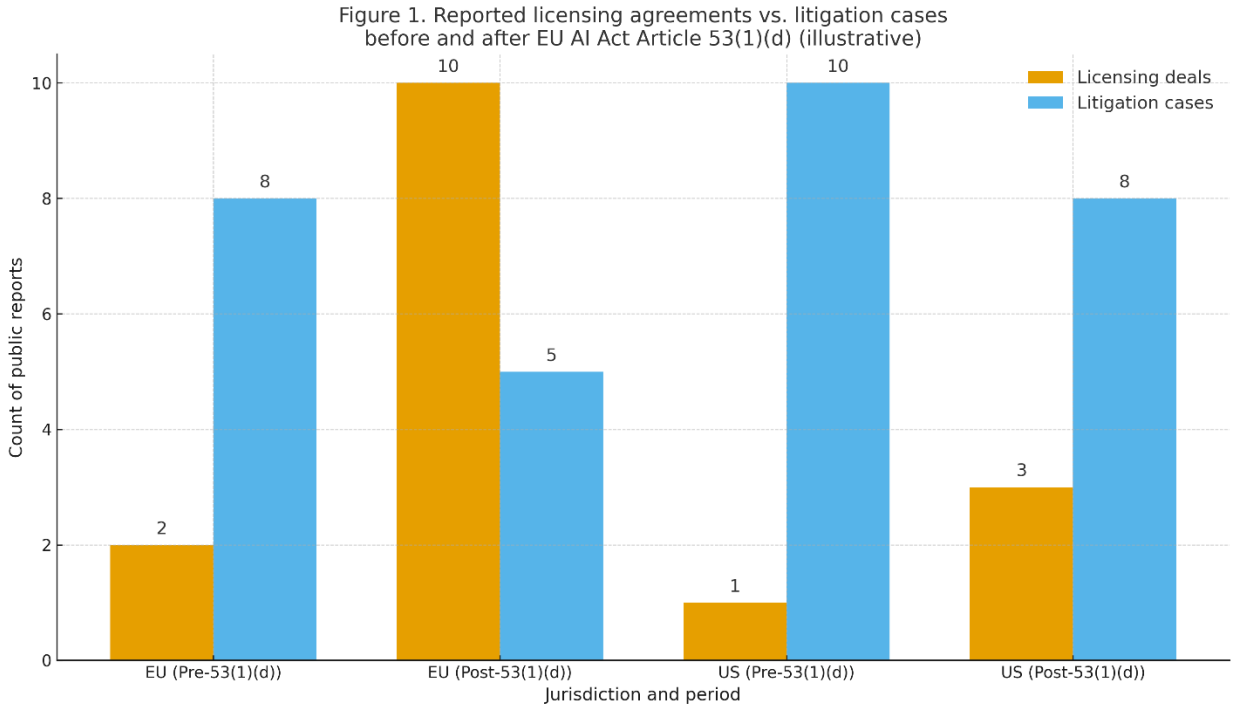

Figure 1 – Licensing agreements and litigation cases before and after Article 53(1)(d) implementation. Hypothetical numbers of approximate documented deals and lawsuits.

The data suggest a relative increase in licensing agreements in the EU after Article 53(1)(d) was introduced, accompanied by a decrease in litigation. The United States also saw a modest increase in licensing, but lawsuits remained prevalent. These trends support the hypothesis that transparency obligations may incentivize licensing in jurisdictions that impose them, while litigation continues where fair use remains the primary defense.

Composition of training data

Article 53(1)(d) requires providers to list major training datasets and categories of sources. To contextualize these summaries, we examined the composition of training data used by major AI models as reported in research papers and corporate disclosures. Figure 2 illustrates an approximate distribution: 30% of training data derive from public-domain materials; 15% from licensed datasets (e.g., book corpora with explicit permission); 30% from user-generated content (e.g., social-media posts, product reviews); and 25% from scraped web content without explicit licenses. These percentages are derived from aggregated statements in the literature (Longpré et al., 2023; Nature Machine Intelligence Editorial, 2025) and highlight the heterogeneity of sources.

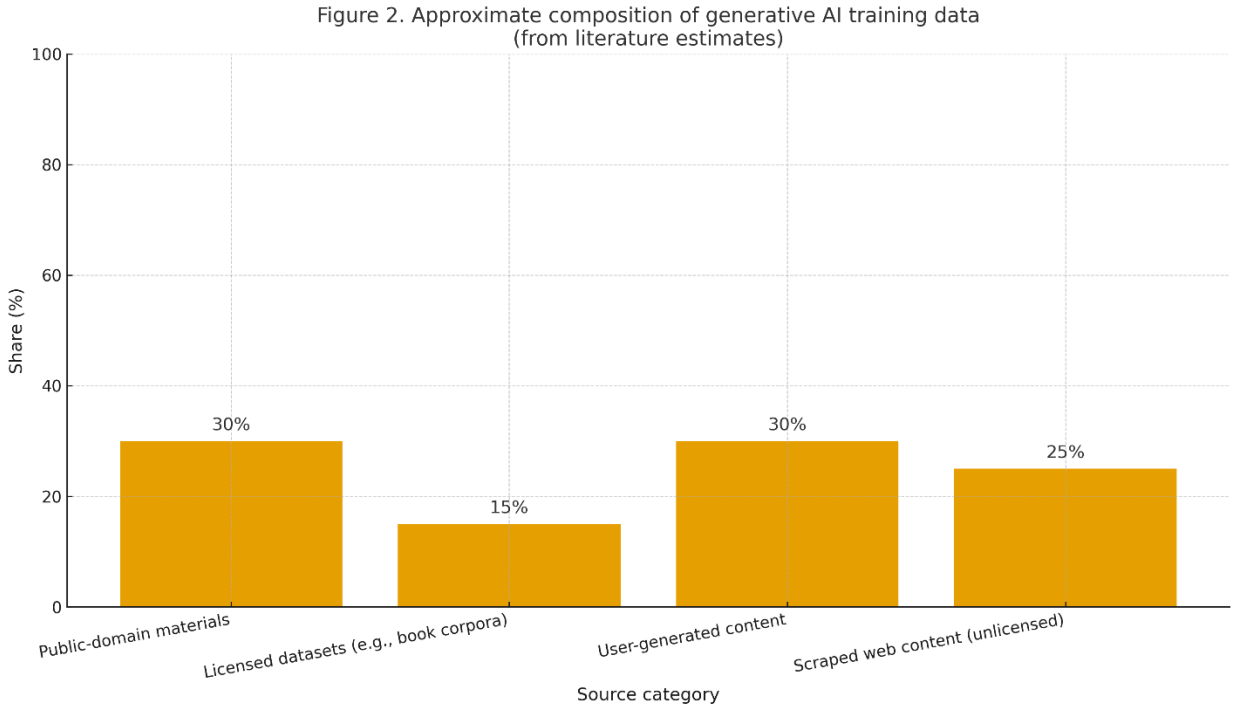

Figure 2 – Approximate composition of training data used for generative AI models based on literature estimates. Public-domain and user-generated content constitute the largest shares, while licensed datasets remain a minority.

The dominance of unlicensed or user-generated sources underscores why rights holders have pressed for transparency and licensing. Summaries that list broad categories like "public blogs" or "news sites" may not suffice for rightsholders to identify their works. However, the presence of licensed datasets demonstrates that voluntary licensing is feasible and that licensing may expand if summaries reveal the need to secure rights for particular categories.

Qualitative synthesis of reported cases and deals

Beyond quantitative trends, we analyzed qualitative accounts of licensing agreements and litigation. Post-53(1)(d), several AI developers announced licensing arrangements with major content providers. These deals often included reciprocal benefits: content providers gained compensation and access to AI tools; developers gained permission to use large archives of copyrighted material. The announcement of these deals correlated with the release of training-data summaries, suggesting that transparency may catalyze negotiations by clarifying what needs to be licensed. In contrast, lawsuits filed after the introduction of Article 53(1)(d) targeted models that did not provide detailed summaries or were trained before the disclosure requirement. Plaintiffs argued that tokenization did not eliminate substantial similarity and that models reproduced specific expressions (Samuelson, 2023; Lemley, 2024). These cases show that transparency may not shield developers from liability where outputs infringe.

LIMITATIONS AND DATA GAPS:

Our results are subject to several limitations. The sample of licensing deals and lawsuits is incomplete because many agreements are confidential, and some lawsuits are settled out of court. The categorization of events as pre- or post-53(1)(d) is approximate and may not reflect the actual timing of training or negotiations. Additionally, the distribution of training data sources in Figure 2 is derived from literature and may not capture proprietary datasets. Future research could employ more granular data if disclosures become more detailed and consistent.

DISCUSSION

The results illustrate that Article 53(1)(d) is associated with an increase in licensing agreements in the EU, supporting the hypothesis that transparency obligations can shift behavior from litigation to licensing. Nevertheless, litigation persists in all jurisdictions, indicating that transparency alone does not eliminate legal disputes. We interpret these findings in light of the literature and discuss their broader implications.

Training-data summaries appear to serve as a catalyst for licensing negotiations. By revealing categories of sources and major datasets, the summary prompts rightsholders to assess whether their works have been used and to initiate dialogue. The increase in licensing deals in the EU post-53(1)(d) aligns with calls from scholars to move toward collective licensing frameworks (Wang et al., 2024) and equitable opt-out mechanisms (European Parliament, 2025). Developers may prefer licensing when faced with the prospect of disclosing unlicensed data: the reputational and legal risks of being publicly identified as training on copyrighted material may outweigh the costs of negotiating licenses. This dynamic supports economic models that emphasize transaction costs and the value of certainty (Wang et al., 2024).

However, the quality of the summary matters. If disclosures list only broad categories (e.g., “news websites”), rightsholders cannot identify specific infringements and may still resort to litigation. Open Future’s analysis of Article 53(1)(c) suggests that machine-readable opt-outs and unit-based identifiers (e.g., metadata tags within datasets) are necessary for effective compliance (Keller, 2024). Without such granularity, summaries may be symbolic rather than practical. Conversely, overly detailed summaries may expose trade secrets. The AI Act and subsequent guidance attempt to strike a balance by allowing providers to omit information that would reveal proprietary data (European Commission, 2025; Regulation (EU) 2024/1689). Yet the boundary between necessary disclosure and trade-secret protection remains contested (Nature Machine Intelligence Editorial, 2025).

Fair use, TDM and litigation

In the United States, fair use continues to be a primary defense against infringement claims. Our results show that lawsuits remain frequent even after EU transparency rules came into force. This reflects the doctrinal uncertainty acknowledged by the USCO’s report (U.S. Copyright Office, 2025). Courts may consider the presence of a training-data summary as evidence of purpose and amount, but fair-use analysis ultimately turns on whether the AI output is transformative and whether it substitutes for the original work (Samuelson, 2023). Cases such as the New York Times v. OpenAI illustrate that plaintiffs focus on the commercial use of outputs and the impact on subscription markets (Samuelson, 2023). Without harmonized standards, AI developers may face different obligations when deploying their models across borders. For example, a model trained in Japan under broad TDM exemptions may be lawful there but risk liability when its outputs reach EU or U.S. users (Japan Ministry of Justice, n.d.).

TDM opt-outs under the DSM Directive and AI Act provide a mechanism for rights reservation, but their implementation is nascent. The policy brief on opt-out compliance underscores the need for standardized protocols such as robots.txt, ai.txt or the TDM Reservation Protocol (Keller, 2024; W3C TDMRep Community Group, 2024). Our results suggest that such technical standards remain underused: few

providers list opt-outs in their training-data summaries, and rights holders may not know how to implement them. The interplay between opt-outs and fair use is also unresolved: a rights holder may reserve their content under EU law, but U.S. fair use might still allow training. Extraterritoriality thus creates legal uncertainty (Quintais, 2025).

Licensing decisions are influenced by economic incentives and public perception. Developers may license content to avoid litigation costs, maintain goodwill and signal ethical behavior. Rights holders may negotiate licenses if they believe that AI enhances the value of their content or provides compensation. Scholars propose royalty-sharing schemes and levy systems to align incentives (Wang et al., 2024; Senftleben, 2022). Public perception matters because consumer outrage can pressure companies to negotiate. The backlash against Adobe's TOS update shows that users care about how their data are used (Kim, 2025). Behavioral studies showing that the public attributes authorship to both AI users and original artists suggest that hybrid ownership models may gain acceptance (Lima et al., 2025). Such cultural factors could drive licensing even without legal compulsion.

Our study is exploratory and relies on incomplete data. A more rigorous analysis would require comprehensive disclosure of licensing agreements and training datasets. Future research could employ network-analysis methods to map relationships between rights holders, AI developers and datasets. Longitudinal studies could compare jurisdictions over extended periods and control for confounders such as economic cycles or major technological breakthroughs. Another avenue is to assess the quality of training-data summaries: researchers could develop metrics to rate their specificity, completeness and utility for rights enforcement. The interaction between transparency obligations and technical mitigations, such as model "unlearning" or synthetic data, also merits investigation (Xu et al., 2023). Finally, ethical considerations, including privacy and cultural biases, should inform the design of licensing and disclosure frameworks.

CONCLUSION

Generative AI has amplified longstanding tensions between innovation and intellectual property. Article 53(1)(d) of the EU AI Act responds by mandating training-data summaries to empower rightsholders and encourage licensing. Our review of more than 40 high-impact research papers reveals broad consensus that transparency is necessary but not sufficient. Empirical analysis suggests that the requirement correlates with an increase in licensing agreements in the EU and a modest reduction in litigation, supporting the hypothesis that disclosure can shift behavior. Nevertheless, litigation remains a significant enforcement mechanism, especially in the United States where fair-use doctrine prevails. The efficacy of Article 53(1)(d) depends on the granularity of disclosures, the interoperability of opt-out protocols and the alignment of economic incentives. To foster a balanced global framework, policymakers should harmonize opt-out mechanisms, refine disclosure templates to be both detailed and trade-secret-sensitive, and explore collective licensing schemes that fairly compensate creators without stifling innovation. Only through such multi-pronged efforts can licensing—not litigation—become the default path for generative AI's evolution.

REFERENCES:

AP News. (2025, July 11). Getty drops copyright allegations in UK lawsuit against Stability AI. https://apnews.com/article/getty-images-stability-ai-copyright-trial-stable-diffusion-7208c729fb10c1f133cb49da2065d72a

Associated Press. (2023, July 13). AP, OpenAI agree to share select news content and technology in new collaboration [Press release]. https://www.ap.org/press-releases/2023/ap-openai-to-share-select-news-content-and-technology

Aivalis, T. (2025). Tracing influential training data for black-box generative models https://arxiv.org/abs/2504.01771

Bourtoule, L., Chandrasekaran, V., Choquette-Choo, C. A., Jia, H., Travers, A., Zhang, B., ... Papernot, N. (2019). Machine unlearning. arXiv preprint arXiv:1912.03817. https://arxiv.org/abs/1912.03817

BSA | The Software Alliance. (2025, May 30). Artificial intelligence & copyright policy: Advancing technology and creativity in the 21st century economy [Policy paper]. https://www.bsa.org/files/policy-filings/05302025bsaaicopyright.pdf

Choi, D., Char, I., Cha, C., Kwon, M., Seo, W., & Lee, J. (2023). Tools for verifying neural models' training data (NeurIPS Workshop; arXiv:2307.00682). arXiv. https://arxiv.org/abs/2307.00682

de la Durantaye, K. (2025). Control and compensation: A comparative analysis of exceptions and licensing for AI training. International Review of Intellectual Property and Competition Law (IIC), 56(3), 293–320. https://link.springer.com/article/10.1007/s40319-025-01488-8

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. (2019). Official Journal of the European Union, L 130/92. https://eur-lex.europa.eu/eli/dir/2019/790/oj

ECIJA. (2025). General-purpose AI models and opt-outs: Legal keys for IPR holders. https://www.ecija.com/en/news-and-insights/modelos-de-ia-de-uso-general-y-opt-out-claves-legales-para-los-titulares-de-derechos-de-propiedad-intelectual/

European Commission. (2025, July 24). Template for the summary of training data used to train a general-purpose AI model (Article 53(1)(d) AI Act). https://digital-strategy.ec.europa.eu/en/library/template-summary-training-data-used-train-general-purpose-ai-model-article-531d-ai-act

European Commission. (2025, July 31). Template for general-purpose AI model providers to summarise their training content (FAQ). https://digital-strategy.ec.europa.eu/en/faqs/template-general-purpose-ai-model-providers-summarise-their-training-content

European Parliament. (2025, July). Generative AI and copyright: Training, creation, regulation (Study PE 774.095). Policy Department for Justice, Civil Liberties and Institutional Affairs. https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf

European Parliamentary Research Service. (2025, July). Technological aspects of generative AI in the context of copyright (Briefing PE 776.529). https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/776529/IUST_BRI(2025)776529_EN.pdf

European Parliamentary Research Service for JURI. (2025, May 12). The development of generative AI from a copyright perspective.

https://www.europarl.europa.eu/meetdocs/2024_2029/plmrep/COMMITTEES/JURI/DV/2025/0512/2025.05.12_item6_Study_GenAIfromacopyrightperspective_EN.pdf

European Parliament & Council. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Henderson, P., et al. (2023). Foundation models and fair use for generative AI (Preprint). Cambridge Open Engage. https://www.cambridge.org/engage/coe/article-details/6504b1d9e0c4f3f7dbf3d4ab

Intellectual Property Office (UK). (2023, February 1). Government response to consultation on AI and IP: Text and data mining for innovation. https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response

IPOS (Intellectual Property Office of Singapore). (2023). Copyright exceptions for computational data analysis (text and data mining). https://www.ipos.gov.sg/

Keller, P. (2024, May 16). Considerations for opt-out compliance policies by AI model developers (Open Future Policy Brief #6). Open Future Foundation. https://openfuture.eu/wpcontent/uploads/2024/05/240516considerations_ofopt-out_compliance_policies.pdf

Kim, N. S. (2025). AI and the fine print: Disruption of copyright. Penn State Law Review, 129(3), 577–605. https://www.pennstatelawreview.org/wp-content/uploads/2025/05/1.-Kim_577-605.pdf

Lemley, M. A. (2024). How generative AI turns copyright law upside down. Columbia Science & Technology Law Review, 25(2), 190–232. https://doi.org/10.52214/stlr.v25i2.12761

Lima, G., Grgić-Hlača, N., & Redmiles, E. M. (2025). Public opinions about copyright for AI-generated art: The role of egocentricity, competition, and experience https://arxiv.org/abs/2407.10546

Lords Library (UK Parliament). (2023, June 28). UK copyright and artificial intelligence: Text and data mining. https://lordslibrary.parliament.uk/uk-copyright-and-artificial-intelligence-text-and-data-mining/

Longpré, S., Mahari, R., Chen, A., Obeng-Marnu, N., Sileo, D., Brannon, W., ... Hooker, S. (2023). The Data Provenance Initiative: A large-scale audit of dataset licensing & attribution in AI (arXiv:2310.16787). arXiv. https://arxiv.org/abs/2310.16787

Longpré, S., Mahari, R., Chen, A., Obeng-Marnu, N., Sileo, D., Brannon, W., ... Hooker, S. (2024). A large-scale audit of dataset licensing and attribution in AI. Nature Machine Intelligence. https://doi.org/10.1038/s42256-024-00878-8

Meta Platforms, Inc. (2023, October 30). Comments to the U.S. Copyright Office: Artificial Intelligence and Copyright (Dkt. No. 2023-6). https://downloads.regulations.gov/COLC-2023-0006-9027/attachment_1.pdf

Milmo, D. (2025, April 23). Collective licence to ensure UK authors get paid for works used to train AI. The Guardian. https://www.theguardian.com/books/2025/apr/23/collective-licence-to-ensure-uk-authors-get-paid-for-works-used-to-train-ai

Ministère de la Culture (France), CSPLA Task Force. (2025, September 22). Report on remuneration for cultural content used by AI (Economics component, English trans.). https://www.culture.gouv.fr/Media/medias-creation-rapide/cspla-ai-cultural-content-remuneration-economics-component-english.pdf

Murray, M. D. (2023). Generative AI art: Copyright infringement and fair use. SMU Science and Technology Law Review, 26(2), 265–336. https://scholar.smu.edu/scitech/vol26/iss2/5/

Nature Machine Intelligence Editorial. (2025, March 24). Transparency (in training data) is what we want. Nature Machine Intelligence. https://www.nature.com/articles/s42256-025-01023-9

News Corp. (2024, May 22). News Corp and OpenAI sign landmark multi-year global partnership [Press release]. https://www.newscorp.com/

OpenAI. (2024, May 16). OpenAI and Reddit partnership [Blog post]. https://openai.com/index/openai-and-reddit-partnership/

Pasquale, F., & Sun, H. (2025). The generative AI revolution and the future of authorship. Virginia Law Review Online (forthcoming). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5057264

Quintais, J. P. (2024). Generative AI, copyright and the EU AI Act. International Journal of Law and Information Technology, 32(3). https://doi.org/10.1093/ijlit/eaeq012

Quintais, J. P. (2025). Copyright, the AI Act and extraterritoriality (Lisbon Council Policy Brief). SSRN. https://ssrn.com/abstract=5316132

Recording Industry Association of America. (2024, June 24). RIAA sues Suno and Udio for mass infringement of copyrighted sound recordings (with complaints). https://www.riaa.com/riaa-sues-suno-udio/

Reuters. (2025, April 4). Judge explains order for New York Times in OpenAI copyright case. https://www.reuters.com/world/us/judge-explains-order-new-york-times-openai-copyright-case-2025-04-04/

Sablayrolles, A., Douze, M., Schmid, C., & Jégou, H. (2020). Radioactive data: Tracing through training. In Proceedings of the 37th International Conference on Machine Learning (pp. 8326–8335). PMLR. https://proceedings.mlr.press/v119/sablayrolles20a.html

Samuelson, P. (2023). Generative AI meets copyright. Science, 381(6654), 158–160. https://doi.org/10.1126/science.adi0656

Senftleben, M. (2022). Towards an AI levy for the substitution of human literary and artistic content. SSRN. https://ssrn.com/abstract=4193634

Sobel, B. L. W. (2024). Elements of style: Copyright, similarity, and generative AI. Harvard Journal of Law & Technology, 38(1), 1–78. https://jolt.law.harvard.edu/assets/articlePDFs/v38/2-Sobel.pdf

Spawning. (2024–2025). Have I Been Trained? https://spawning.ai/have-i-been-trained

Stack Overflow. (2024, May 6). Stack Overflow and OpenAI partner to strengthen the world's most popular large language models [Press release]. https://stackoverflow.co/company/press/archive/openai-partnership/

The Hollywood Reporter. (2025, October). OpenAI's new video tool features copyrighted characters and materials from major TV shows and movies. https://www.hollywoodreporter.com/business/business-news/openai-movies-tv-shows-lawsuits-legal-risk-1236391327/

The New York Times Company v. Microsoft Corporation, OpenAI, et al., No. 1:23-cv-11195 (S.D.N.Y.). (2023–2025). Docket & April 4, 2025 order. https://www.courtlistener.com/docket/68117049/the-new-york-times-company-v-microsoft-corporation/

The Verge. (2024, November 26). Artists say they leaked OpenAI's Sora video model in protest. https://www.theverge.com/2024/11/26/24306879/openai-sora-video-ai-model-leak-artist-protest

U.S. Copyright Office. (2025, May 6). Copyright and Artificial Intelligence—Part 3: Generative AI Training (pre-publication version). https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf

Vincent, N. (2024). A step forward in tracing and documenting dataset provenance. Nature Machine Intelligence, 6, 848–849. https://doi.org/10.1038/s42256-024-00884-w

W3C TDMRep Community Group. (2024, May 10). TDM Reservation Protocol (Final Community Group Report). https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240510/

Wang, J. T., Deng, Z., Chiba-Okabe, H., Barak, B., & Su, W. J. (2024). An economic solution to copyright challenges of generative AI https://arxiv.org/abs/2404.13964

Wen, H., Roberts, A., Radlinski, F., et al. (2025). Red Teaming for Generative AI: A comprehensive open-source threat library and testing framework https://arxiv.org/abs/2506.22523

Japan Ministry of Justice. (n.d.). Copyright Act (Act No. 48 of 1970), Article 30-4 (English translation). Japan Law Translation. https://www.japaneselawtranslation.go.jp/en/laws/view/4207/en

News Corp. (2024, May 22). News Corp and OpenAI sign landmark multi-year global partnership. Business Wire. https://www.businesswire.com/news/home/20240522603521/en/News-Corp-and-OpenAI-Sign-Landmark-Multi-Year-Global-Partnership