Legal Research & Analysis

Innovative

Research Article

Copyright Enforcement in the Age of Generative AI

Aranya Nath[1,*] and Gautami Chakravarty[2]

[1] DPIIT IPR Chair, Hidayatullah National Law University, Naya Raipur, Chhattisgarh 492099, India
[2] Amrita International School of Law, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu 641112, India; c_gautami@cb.amrita.edu
* Correspondence: aranya.nath@hnlu.ac.in

Abstract: Generative artificial intelligence (AI) raises significant copyright enforcement issues regarding source acquisition, training data retention, optimization, and output. Current study examines IP rights enforcement via mixed doctrinal and descriptive and documentary methodology where opaque technology intersects with undefined legal standards. The research reviews 97 recorded generative AI cases from 2022 to 2026, including 13 judgments or procedural entries and 13 legislative or other government documents. A total of 132 recorded instances were counted, and duplicate and irrelevant entries were excluded. The proceedings register reflects 79 U.S. cases (81.4% of the total), of which 51 were filed in the Southern District of New York (SDNY). Recorded filings increased from two in 2022 to 37 in the incomplete observation period of 2026, while 62 proceedings show an active public-status label. The largest single uncombined enforcement category concerns training-use allegations, with 32 entries. It neither establish the prevalence of infringement nor evaluate the success of related litigation. Doctrinal analysis reveals that certain evidentiary and remedial issues arise from the reproduction right, training exceptions, rights reservations, human authorship, and copyright management information, among others. The cited judicial developments underscore the importance of distinguishing between allegations and holdings, the source of the training material and subsequent training, and interim versus final decisions. The paper presents a framework that integrates identification and traceable acquisition of rights, their disclosure and legal defence, and the proportionality of the relief sought. It supports documented acquisition practices, interoperable rights reservations, practical licensing pathways, and access to courts while maintaining confidentiality. The framework addresses enforcement barriers and structures responses in a controlled manner. Further research is needed to verify dockets and track outcomes to substantiate stronger claims.

Keywords: generative artificial intelligence; copyright enforcement; fair use; text and data mining; AI training; litigation; transparency; licensing

Citation: Aranya, Nath and Gautami Chakravarty. 2026. Copyright Enforcement in the Age of Generative AI. *Legal Research & Analysis* 4(3), 34-42. https://doi.org/10.69971/lra.4.3.2026.209

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license https://creativecommons.org/licenses/by/4.0/.

1. Introduction

Generative AI creates text, images, sounds and other content by using digitized information as parameters. Developing these systems usually involves (a) obtaining and organizing content, (b) training and fine-tuning the system, (c) making the system publicly available, and (d) allowing users to generate content using the system. Each of these steps is distinct. Obtaining the content may trigger the reproduction right, regardless of whether the finished system produces any recognisable content (Saw and Tan 2025). Distributing a system raises different questions of protectable expression and content. Treating every aspect of a generative AI system as a single, unified act of infringement prevents courts from determining which rights have been violated and applies the wrong remedies (Wu 2024).

This research focuses on cases where evidence of infringement is lacking rather than on the copyrightability of machine-generated content. A right may exist in theory while evidence establishing copying or unauthorised access or distribution is locked inside a developer's infrastructure. On the other hand, the appearance of a particular style does not mean that protected expression has been unlawfully copied. Adjudication in a relevant jurisdiction must establish a link between a specific work and a specific rights holder and identify an exception and a remedy that can be provided (Guildea 2026). The study considers a restricted register of copyright-related proceedings and selected legal developments up to and including 12 October 2026. The analysis does not assume that all entries in the register pertain to infringement that has been established, or that all relate to training. The classification systems and comparative law used in the entry pertain to en-

forcement and other issues without resolving factual disputes. A single entry may pertain to multiple works, claims and/or stages under various, and possibly conflicting, laws of different jurisdictions. This research aims to map recorded proceedings by year, jurisdiction, protected-work classification, enforcement question, and reported public status. The second aim is to compare the assessments of courts and regulators regarding acquisition, training and generation in the four selected legal frameworks (Buick 2025). The third aim is to assess the responses that might reduce information asymmetry, safeguard legitimate research and innovation interests, and protect confidential technical systems (Zhang 2025).

The questions are legal and empirical. Where are the most documented conflicts? Which rights or what facts are most often in dispute? What distinguishes a claim from a judgment or settlement? Can exceptions, notices, and disclosure help owners figure out how their works were used? What is the best combination of private licensing, judicial process and public oversight? Will that change the ability to easily challenge the unwanted use of their works?

This work is doctrinal and observed features of the process for enforcement. Numbers can't show that infringement is increasing. Filing patterns are a function of a myriad of factors including access to and incentives for counsel, how the case attracted publicity, the preferences of the claimants and forums, and how the records were created. A single decision may not resolve the legality of the technology in different jurisdictions. This work therefore locates the most disputed areas and use doctrinal comparisons to identify the questions that determine the legal consequences.

Figure 1 shows this work in the larger context of the institutional organization of copyright administration. Registries, courts and public guidance is important when technologically complex, creative activities are administered. These institutions identify genuine ownership and copying questions and distinguish them from economic uncertainties, so that standards and practices help the parties resolve their concerns without the need to litigate.



Figure 1. James Madison Memorial Building, Library of Congress, Washington, D.C. This 2007 photo by Carol M. Highsmith highlights the discussion within public domain copyright administration instead of depicting a disputed AI output. It shows ongoing relevance of available guidance and procedures for rights administration, registration, and enforcement.

2. Background

The training stage demonstrates that access to information and acts for copyright owners are distinct in law. Developers generally don't seek permission to access publicly available works. They may do without legal obligation to reproduce, store or use the information for all subsequent purposes. Training pipelines may retain intermediate copies, preserve source material or ingest third-party collections, based on different contractual conditions. Not all pipelines respect the same copyright treatment. The fact that a model is subsequently offered to consumers cannot be used to conclude how pipelines respect copyright treatments (Wu 2024).

Training-related reproductions are assessed in the United States analogous to the statutory four-factor fair-use inquiry. This involves an examination of the purpose of the reproduction, the nature of the work, how much of the work is reproduced, and the potential effect of the reproduction on the market for the work. For this analysis, commercial actor interests typically align with the interests of educational institutions and scholars. While the EU doesn't include general exceptions and limitations, it does include numerous exceptions for text and data mining (de la Durantaye 2025). Systems may serve commercial interests in both systems doesn't mean the applicable defence and exceptions are the same (Manteghi 2024).

Three questions warrant separate proof: whether the source was obtained lawfully; the number of copies made or retained; and the legally significant uses during the training. A claimed exception for analytical processing need not dispose of independent allegations concerning unauthorized acquisition. Similarly, copying in the course of a training session done not demonstrate an infringing substitution with respect to that output.

Flexible exceptions for technical development are one possible approach. A rights-clearance mechanism could also be devised. A 'license on demand' system could, in theory, fill the same function as an exception by protecting the position of creators. There must, however, be a consideration of practical rights-reservation mechanisms (Tyagi 2024).

Copyright for generated material should be treated separately from the legality of model development. For AI-assisted work, elements authored by a human may be protectable if a sufficient level of original expression is present. As such, content generated by a computer may be treated differently regarding registration and title. The creative input matters more than the presence of a prompt, software or a human operator. For creative works, originality, selection and arrangement are treated differently (Mazzi 2024).

The contrary question is whether a generated response reproduces someone else's protected work. Under some conditions, diffusion models can emit fragments of material associated with the training examples. The ability to extract examples from a model

in this way establishes a risk mechanism, not a court-actionable infringement claims with an identified defendant in an identified proceeding. For an infringement claim, evidence of the protected expression, the relevant act, the elements of the cause of action and/or defense, and the relevant legal standards are still required (Carlini et al. 2023). The similarity of the outputs as well as the degree of copying and the relationship of the works to the protected works is considered. Shared ideas and style do not usually satisfy the test. Evidence that the outputs displace licensed access to the original may be evidence of harm, or may support a fair use claim. Whether the evidence is legally persuasive depends on the cause of action and the jurisdiction.

Attribution and copyright management information is another aspect. Removing copyright-related metadata from a work acquired for training may prompt claims that are not related to the expressiveness of the work. Legal analysis should distinguish between the author, the source of the work, memorizing the work, reproducing the work, and the rights related to the work. Otherwise, evidence concerning one aspect may be wrongly construed to be about another aspect.

One unevenness in enforcement is that while rights holders generally can see publicly released outputs, they can't see records of source acquisitions, preprocessing, or model optimisation. A pleading may identify works and possibly suspect training practices without showing that the defendant accessed a particular copy. Courts must avoid demands for proof that is impossible to provide at the stage of a lawsuit and to prevent overbroad discovery. Detailed consideration of pleadings, the parties and the burdens of proof is critical to accountability (Buick 2025).

Examples of proposed disclosures include training summaries, source records, rights reservations, and compliance policies. They have various purposes. A general summary could indicate the training sectors without allowing an author to identify a particular work. There is a risk that machine-readable notices will not work if intermediaries remove identifiers. Collective licensing can facilitate uses of works and lower transaction costs for works in well-organised repertoires, while failing to consider creators where their ownership or representation is uncertain (Peukert and Castets-Renard 2025).

There are limits to what can be accomplished through technical governance. An acquisition catalogue, for example, is only as trustworthy as the acquisition process that created it. Attribution tools can create false accusations and false reassurances. Developers withhold information that exposes security vulnerabilities, personally sensitive information, or private implementation details. Independent review, appropriate confidentiality controls and proportionate disclosure are more trustworthy than reliance on the general public to detect and act upon information in training artefacts (Pasetti et al. 2025).

The research gap is not a lack of fair use, authorship or licensing individually. It is the failure to integrate legal rights with appropriate evidentiary access, and varying types of remedies. This study places bounded proceeding frequencies beside comparative doctrinal analyses to identify where enforcement gaps exist and recommend how institutions should respond to such gaps, without confusing the documentary register with a representative measure of legal wrongdoing.

3. Comparative Legal and Enforcement Framework

3.1 United States: Exclusive Rights, Fair Use and CMI

United States copyright law offers several avenues for disputes involving generative AI. Sections 106 and 106A of Title 17 of the U.S. Code outline the exclusive rights of copyright holders, such as the right to reproduce the work or to make a derivative work. Section 107 codifies fair use and identifies four nonexclusive factors. Claimants must prove their ownership of the copyrighted work and defendants' unauthorized use. Defendants may raise a defense if applicable. Different theories should be treated differently despite similar facts of the case. For example, in the case of copyright management information, codified under section 1202, the same principles should apply (Ginsburg 2025).

The recorded 2025 decisions in *Bartz v. Anthropic*[1] and *Kadrey v. Meta*[2] illustrate why the concrete evidentiary record matters. Both concern training and fair use but differ on acquisition and market evidence. *Bartz* distinguishes training from acquiring and retaining pirated copies. In *Kadrey*, the entry highlights the plaintiffs' failure to establish meaningful market harm on the record before the court. Neither authorizes all copyrighted training uses (Saw and Tan 2025).

A separate inquiry is necessary for questions involving market substitution and output stage reproduction. Using an item as work for training doesn't ensure that aspects of the item become protected in the system's responses. For instance, a claim concerning the removal of identifying information may succeed or fail based on the provisions of section 1202, regardless of whether an independently infringing passage is generated.

Copyright registration guidance further distinguishes material determined by a human creator from material determined by a machine. In effect, this means claimants have to show exactly which part of a copyrighted work they own. Status of a proceeding, for example an appeal or an interim order, doesn't by itself indicate that a disputed training practice has finally been held lawful or unlawful (Mazzi 2024).

3.2 European Union and United Kingdom: TDM, Disclosure and Rights

EU training disputes follow a specific text-and-data-mining framework. Article 4 of Directive (EU) 2019/790 allows mining of legally accessible works, subject to valid rights reservations. Regarding reservations, acquisition systems should be able to recognize conditions attached to accessible material (Tyagi 2024). This exception, different from the American fair use, still requires consideration of authorisation and use (de la Durantaye 2025).

Article 53(1)(c) of Regulation (EU) 2024/1689 requires providers of general-purpose AI models to adopt a copyright compliance policy. Article 53(1)(d) requires a sufficiently detailed public summary of the training content. Commitments to policy adoption and public summaries of training content are governance obligations. They do not represent a judicial determination that specific

---

1 *Bartz et al. v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal. 2025), https://www.govinfo.gov/app/details/USCOURTS-cand-3_24-cv-05417 (last visited October 19, 2026).

2 *Kadrey et al. v. Meta Platforms, Inc.*, No. 3:23-cv-03417, Document 598 (N.D. Cal. 2025), https://law.justia.com/cases/federal/district-courts/california/candce/3:2023cv03417/415175/598/ (last visited October 19, 2026).

works were lawfully acquired. The usefulness of commitments and policy statements depends on the scope, detail and the procedures in place to ensure compliance.

During the output phase, several issues relating to recognized reproduction, communication to the public and use of protected expression require the application of copyright doctrines to the identified conduct. The existence of a broad regulatory transparency duty would not resolve the infringement standard for any of the model outputs, nor would it act as proof of access and copying; such elements are material.

The Copyright, Designs and Patents Act 1988 apply in the UK, and so does section 16 of that Act, which lists limited exceptions to copyright. The country also has specific legislation relating to text and data mining. Judgment was handed down in *Getty Images v. Stability AI* (UK) and refers to certain secondary infringement actions concerning model weights and provides an update concerning an appeal. The judgment does not address whether all generative models infringe copyright or whether all uses of photographs in a generative model's training infringe copyright; of course, claim, evidence, remedy, and the stage of the proceedings must all be taken into account when making comparisons.

3.3 India and Comparative Cross-Border Implications

In India, the Copyright Act 1957 identifies exclusive acts in section 14, sets out infringement consequences in section 51 and defines exceptions in section 52. Indian fair dealing is purpose- and condition-restricted, unlike the US fair use. Hence, conclusions drawn regarding a transformative purpose in a US case will not apply to the Indian context. The analysis must consider the protected work and the contested act. It should consider whether a particular exception applies to the case (Sriram 2025).

Indian examples relate to the existence of disputes and preliminary procedural questions, rather than settled positions regarding training of generative AI. One such example is the case of *Asian News International v. OpenAI*.[3] An example of interim-relief litigation and its importance is illustrated in this case, as it shows the gap between an order of interim injunction and a formal determination of liability. The Court granted interim relief in the absence of a decision on liability.

Rights owners may be in different countries than intermediaries who host works. Model developers may be in different countries still. The laws that govern rights and remedies may vary from where the rights owner is to where the intermediary is to where the developer is. Where enforcement is practical varies on a case-by-case basis, considering whether and to what extent a service provider cooperates and if there are sufficient legal tools to gain access to the service provider's internal information (de la Durantaye 2025).

Five criteria guide comparison: Criteria such as consent, exceptions, disclosure, remedies and small creator access show differing legal frameworks and vary in the breadth of rights, including institutional and protection resources.

Table 1 provides samples of various legal entry points. These should not be confused with identical treatment. Cross-border cooperation should preserve national disparities and accommodate asynchronous professional and practice differences, especially in the field of evidence.

Table 1. Comparative Copyright Enforcement Instruments.

Jurisdiction	Key provision	Enforcement focus
India	Copyright Act 1957, §§ 14, 51, 52	Exclusive rights; infringement; statutory exceptions
United States	17 U.S.C. §§ 106, 107, 1202	Exclusive rights; fair use; copyright-management information
European Union	DSM Art. 4; AI Act Art. 53	TDM reservations; GPAI copyright policy and summaries
United Kingdom	CDPA 1988, § 16	Restricted acts and remedies

Note: The comparison identifies different statutory gateways through which training and output disputes may be assessed. It is a doctrinal guide, not a claim that the listed provisions produce identical outcomes.

4. Research Methodology

This study employs both doctrinal analysis and descriptive documentation. It employs no experimental methods, nor does it rely on first-hand accounts. Analysis interprets portions of U.S. copyright law, relevant governmental documents, and legal scholarship. The other describes a structured register of select U.S. patent office proceedings involving generative A.I. Collectively, these lines of analysis identify both legal and procedural challenges associated with the registered office actions. This analysis is presented separately from any frequency data associated with office action filings.

The structured register describes 97 generative A.I. proceedings. A separate register describes 13 judicial events. The legal frameworks register describes 13 legislative and/or governmental frameworks. They focus different events. A proceeding describes a single unit of contention, and a judicial event describes a proceeding or case at a particular stage of the litigation process. A legal framework unit describes a legislative or governmental framework. The three registers are not combined to describe independent cases or proceedings. "Proceeding" means one entry classified as one type of litigation or formal enforcement in the register, regardless of the number of claims, causes of action, or other legal theories or remedies. In this analysis, "inclusion" also does not mean the litigation or formal enforcement in question meets the requirements of international law. It means the proceeding is in the official record.

---

[3] *Asian News International v. OpenAI*, Delhi High Court (2026); reported in "ANI v. Open AI: Delhi High Court Refuses Interim Injunction in Landmark AI Copyright Dispute," *India Corporate Law* (2026), https://corporate.cyrilamarchandblogs.com/2026/07/ani-v-open-ai-delhi-high-court-refuses-interim-injunction-in-landmark-ai-copyright-dispute/ (last visited August 29, 2026).

No proceeding in this compilation represents all the proceedings filed in the courts or tribunals in any country, or indicates the number of legal disputes in the world. Most entries lack court case numbers and other identifying court case information. The reasons for omissions present challenges in interpreting this compilation and in verifying records. The focus of the comparison of laws and legal procedures is on the rules pertaining to the training, functioning and the practice of the enforcement of a legal act (Pasetti et al. 2025).

Variables include filing year, jurisdiction, work type, enforcement issue and public status. Variables are complete as recorded, and missing or ambiguous values are not resolved. It is possible to reproduce frequencies. Percentages are calculated using the 97 proceedings, unless otherwise indicated.

With respect to description classes, we take a "no taxonomy" approach. For example, the definitions of musical works and sound recordings overlap. Thus, it might be logical to combine the three in the register. However, doing so would require us to develop and use new classification systems throughout the register. When describing sections of the register that overlapped, we did not merge them. Instead, we accepted a higher degree of redundancy to increase the transparency of the results.

We count judgments and orders separately. For example, if a particular case is appealed, we would count the lower court judgment and the appellate court judgment as separate cases. Likewise, a settlement of a case is a separate case from a final court judgment. If some information regarding a case is provided, we do not make any assumptions regarding the case or the information. Instead, we indicate it is a case that merits further review to determine the full status of the case (Guildea 2026).

This source checks identifiers, categories, completeness, tables and figures. Major limitations include bias in selection, imbalance in country coverage, mutability of procedural statuses, diversity of issue labels and absence of a global baseline. Causal estimates, tests of significance and calculation of possible costs caused are absent in the records. Separating analytical units is central to preventing overstatement and is depicted in Table 2. Units related to it require identity checks and may be combined only after.

Table 2. Documentary Data Sources and Units of Analysis.

Workbook sheet	Records	Analytical use
Court Proceedings	97	Filing year, work type, jurisdiction and procedural status
Judicial Findings	13	Selected decisions and procedural events
Legal Framework	13	Legislation and official guidance for comparison

Note: Each of the three sets of entries has specific functions. Proceedings support requests for descriptive counts. Judicial events are associated with bounded doctrinal interpretation and framework. Legal instruments define the comparative framework.

5. Results and Data Analysis

The proceeding register contains 97 records across five filing years. The annual totals are two in 2022, 12 in 2023, 17 in 2024, 29 in 2025 and 37 through the October 2026 cutoff. These counts sum to the complete register and exhibit an increasing sequence within the observed selection. The last year is not a full calendar year, however, and the count for 2026 cannot be read as a finalised annual total or a reliable forecast of litigation in later months.

Jurisdictionally, 79 entries carry the 'United States' label, or 81.4% of the recorded proceedings. Eighteen entries fall under other national or supranational labels. They are spread across Germany, India, the Netherlands, Italy, Denmark, the European Union, France, Canada, the United Kingdom, China, Brazil and the Czech Republic. These are recorded labels, not necessarily comparable forums, legal systems or individual judicial institutions. Germany contributes five entries, while India and the Netherlands contribute two each; the remaining non-U.S. labels appear once apiece.

U.S. concentration may reflect developer location, market size, litigation incentives, specialist counsel and publicity. Collection practices may also shape it. Without a defined global capture method, the data cannot distinguish these possibilities, and no causal explanation can be inferred from the frequencies alone (Buick 2025).

Figure 2 displays the original annual counts. Its rising bars visualize a concentration of recorded activity rather than the incidence of infringement. They should not be used to imply that courts have increasingly ruled against developers or that every newly filed claim identifies proven unauthorized copying. Interpretation of litigation trends requires subsequent docket-level verification and attention to competing regulatory pathways outside courts (Saw and Tan 2025). Rising counts might also reflect better documentation rather than more underlying unlawful activity.

5.1 Protected-Work Sectors and Alleged Enforcement Issues

Proceedings have been recorded in multiple categories of protected expression. The most common label recorded was "News reporting" (20 instances). There were 11 entries each for the labels "Books" and "Online videos." The labels "Music recordings" and "Photos and images" were recorded 8 times each. The label "Song lyrics" was recorded 4 times. The six categories represented in Figure 3 reflect the most common classification. Similar works may be recorded as "Literary works," "Sound recordings", "Musical works", or "Journalistic works." The classification schemes would likely also be reflected in the works cataloged in the figure.

The field covering enforcement and other questions has a different type of concentration. The largest exact classification is "Training-use allegations" (32 times). The classification "Access control circumvention" is reflected 9 times. So too is the classification "Training and output reproduction." The classification "AI answer reproduction" was recorded 6 times. Other classes reflect a range of distinctions including legal and business matters, rights management, and the copyrightability of the works (Wu 2024).

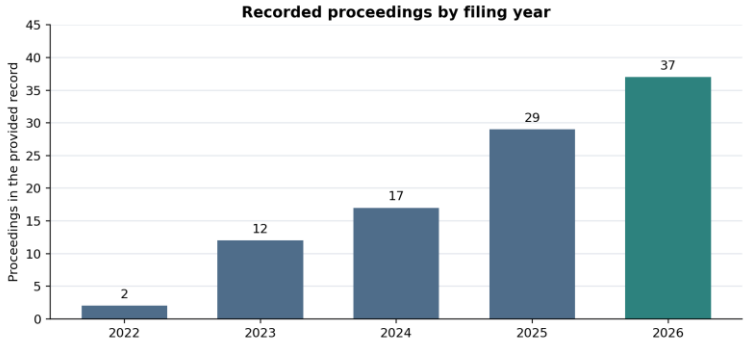


Figure 2. Recorded Generative-AI Copyright Proceedings by Filing Year (2022–2026).

Note: The recorded annual counts increase from 2022 to October 2026. The final year is incomplete, and the collection is selective, so the observed sequence describes this record rather than all copyright proceedings.

We shouldn't automatically equate claims about news reporting with claims about books and recorded music. There may be differences in the rights situations, how content is made accessible, and how it may be replaced. Claims about training copies contain an input-side question, while claims about copied or illustrated contents in responses contain output-side evidence. Circumvention of access control may require other legal conditions, and similarity may not be the only factor for evaluating claims about embedded metadata.

Figure 3 reflects frequency of work rather than prevalence of infringement. There are a limited number of owners exercising control over a significant part of the market, and they may coordinate claims and proceedings. The register does not provide exposure, prevalence or loss figures, and the boundaries of its work categories are also not defined (Buick 2025). Common models can generate distinct disputes; counts cannot measure technical failure or loss.

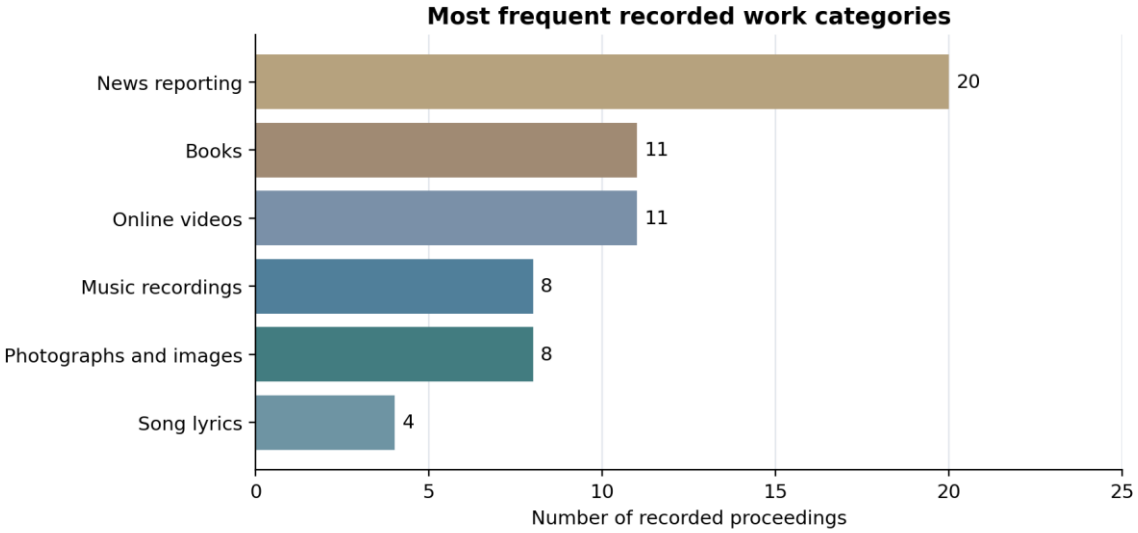


Figure 3. Six Most Frequently Recorded Categories of Protected Works.

Note: Reported news, audiovisual materials (including musical works), books and images are some of the primaries, or 'literal', classifications. These classifications don't, and aren't meant to, quantify infringement, loss, or damage suffered, or estimate the size of the market.

5.2 Judicial Findings and Adjudicative Status

The separate events register lists thirteen judicial and procedural events. Six of these events involve the United States, three, Germany, one each, China, the United Kingdom, Canada and India. Clearly, the entries cannot be construed as 13 different disputes and even less as a sample of the many, if any, judgments rendered. Table 3 shows the sample of the events as recorded.

The United States examples demonstrate a variety of procedures. *Bartz* and *Kadrey* involve a 2025 fair-use summary judgment. The settlement described in the 2026 *Bartz* case[4] was not an unlawful training case. The denial of certiorari in *Thaler v. Perlmutter*[5]

---

4 *Bartz et al. v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal. 2026), https://www.govinfo.gov/app/details/USCOURTS-cand-3_24-cv-05417 (last visited October 19, 2026).

5 *Thaler v. Perlmutter*, No. 23-5233 (D.C. Cir. March 18, 2025), https://law.justia.com/cases/federal/appellate-courts/cadc/23-5233/23-5233-2025-03-18.html (last visited August 29, 2026).

did not address human authorship. Liability as a copyright-management-information provider in *Doe v. GitHub*[6] was at issue. As in Ginsburg (2025), these examples cannot be construed as a win for plaintiff or a loss for defendant.

German entries include text-and-data mining, allegations of reproducing protected song lyrics in generated output, and contract transparency. *GEMA v. OpenAI*[7] is the first proceeding regarding reproduction of song lyrics. The German proceedings and *GEMA v. OpenAI*[8] involve different issues than *Robert Kneschke v. LAION*.[9] The United Kingdom image-model proceeding and Chinese authorship proceeding involve different issues.

These proceedings assist interpretation and foster legal certainty. Judgments depend on several factors including the date of the proceeding and whether an appeal has been taken. Equal weight cannot be given to a first-instance proceeding, a subsequent settlement and an appeal. First-instance proceedings cannot be relied on to establish infringement (Dornis and Lucchi 2025).

Table 3. Geographic distribution of selected judicial events.

Jurisdiction	Recorded event entries
United States	6
Germany	3
China	1
United Kingdom	1
Canada	1
India	1

Note: The case entries are examples of notable judicial developments in six jurisdictions. The same case may be represented more than once. The case entries should not be construed as comparative judicial rankings.

5.3 Procedure, Proof and the Distance between Claims and Relief

Public statuses indicate cases are pending (or 'active') until closed (resolved, settled, or dismissed). Of the 97 cases, 62 are pending, and 35 are closed. The closed cases include adjudication and non-judicial outcomes. 'Active' and 'pending' labels include interim orders of the court, while a 'decided' case indicates the court has ruled, but may provide little information on the winning party and/or the potential for an appeal.

As shown in Figure 4, 'active' cases exceed 'closed' cases. In the absence of complete information, it is unsafe to attribute an enforcement outcome based on the case register. Public labels are temporary (Guildea 2026). Liability in a case may remain undecided, but litigation may be stayed pending the outcome of another case. A case may also be consolidated, dismissed (on or off the record) or settled (privately or commercially).

Best practice in case management is to build evidence incrementally. Disputants must first identify the works at issue and allege ownership. Next, they must allege particular acts and identify the fora within which they occurred. Claims and defences are articulated and adjudicated before similarity and copying are considered. In deciding on infringement, courts examine causation and the appropriate remedy. This procedure minimizes the risk of a general objection based on AI.

Preventive deterring effect balancing (PDEB) posits that the complete disclosure of training corpora may reveal a trade secret, while failing to produce any technical record may deny the rights holder the means to protect their interest. Intermediary measures, including protective orders and sample and restricted inspection by expert(s), may be ordered in advance of disclosure. Courts may order samples of works, including artefacts created by the AI, in cases where adequacy of the sample is challenged by the defendants. Developers are obligated to maintain relevant records and produce them in support of a motion to limit the scope of an order.

5.4 Balancing Remuneration, Innovation and Rights Protection

The central issue is misrepresented as a choice between uncompensated technological progress and unrestricted control by creators. A more refined analysis would identify the interests of creators in licenses and in legal recognition and protection from predatory exploitation. Development in research and technology might promote access, competition and expression. However, when there is monopolization or predatory exploitation, enforcement should address market failure without making an untenable assumption that all industry analysis and strategy is confined to the realm of competition (Zhang 2025).

Where obligations to provide access are reasonably concentrated, affirmative duties may be clearly articulated and agreed upon. Where the opposite is true (i.e., negotiable rights are dispersed), the costs of licensing may not be justified. Transparent and affordable terms for challenge and modification of contracts might facilitate the use of per-copy influence-based royalty systems. Technical similarity of competing works may support an argument for an allocation of copyright ownership (Rai et al. 2025).

---

6 *Doe et al. v. GitHub, Inc. et al.* (N.D. Cal.); see IP Update, https://www.ipupdate.com/tag/doe-et-al-v-github-inc-et-al/ (last visited October 19, 2026).

7 *GEMA v. OpenAI*, Regional Court of Munich I (2025); see Bird & Bird, "Landmark Ruling of the Munich Regional Court (GEMA v OpenAI) on Copyright and AI Training," https://www.twobirds.com/en/insights/2025/landmark-ruling-of-the-munich-regional-court-(gema-v-openai)-on-copyright-and-ai-training (last visited October 19, 2026).

8 *Ibid.*

9 *Robert Kneschke v. LAION e.V.*, Hamburg District Court, case no. 310 O 227/23; see EUIPO, https://www.euipo.europa.eu/en/law/recent-case-law/germany-hamburg-district-court-310-o-22723-laion-v-robert-kneschke (last visited October 19, 2026).

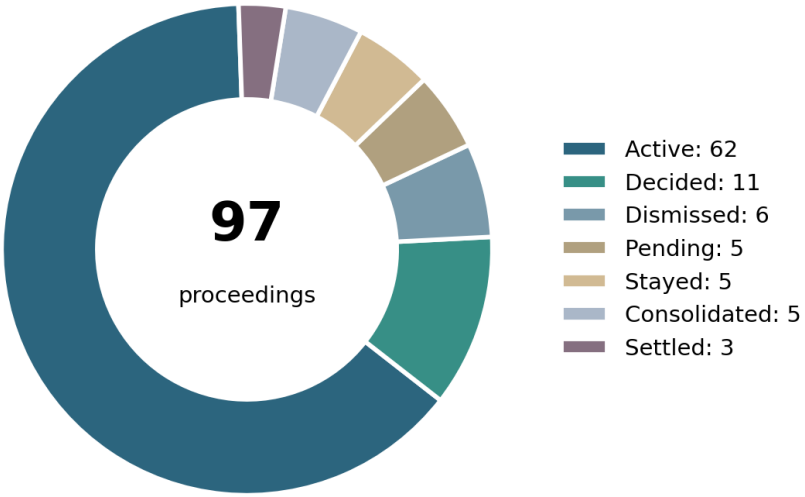


Figure 4. Public Procedural Status of the 97 Recorded Proceedings.

Note: The types of cases that result in classifications represent the majority of cases that are still pending. These cases often illustrate the deficiency of final rulings regarding the merits of claims. Classification statuses are usually temporary, and cases are often modified or amended during litigation.

Opt-out standards, as an intermediate mechanism, can gain greater legal recognition for reservations of rights. They require identifiers, standards, and assurance that notices are not removed by intervening parties. While preservation of notices and logs may assist in accounting for notices, eventually a human will need to assess the logs to determine if notices were properly classified. The procedures describe a framework to justify and account for legal positions, and, in the process, provide evidence relevant to legal positions. These procedures should not provide a means to support a legal conclusion (de la Durantaye 2025).

The procedures have been developed and implemented on a limited basis. User guides and instructions may differ, and relevant links may be missing. Representation may differ from actual values, and assessments to determine policy costs are not possible. Future implementation should require processes and standards to preserve evidence of legal positions and justifications (Haidemariam and Gran 2026).

6. Policy Recommendations

For training data, developers must retain sufficient documentation of the source from which the data was obtained, the kind of data collected, pertinent permits and licenses, and any modifications to the data made by the developers. These records may be kept in a restricted access file. However, a layered approach may allow the public access to a basic description of the records while allowing regulators access to a more detailed description. Records may be made available to a court only upon the order of the court. These records allow developers to demonstrate that they have properly documented the relevant information. The records do not need to be made publicly available (Pasetti et al. 2025).

Substantive rights should allow for permissions and objections. Notices should be automatically transferable and machine-readable. Licensing systems should allow smaller content creators to legally license their works. Not all persons and entities should enter into hundreds or thousands of individual licensing agreements to make their content available lawfully. Some collective or blanket licensing should be permitted where the rights of content creators can be ascertained. Universal adoption of this approach should not be required (Rai et al. 2025).

Procedural reforms to address information asymmetry may include the adoption of a qualified redaction rule for intellectual property claims. Courts may allow complainants to identify (i.e., provide the court with notice of) the works at issue and relevant acquisition channels. To protect the works, courts may limit access to copies of the works to qualified individuals and/or require sampling of the works. A qualified individual may help the court determine whether copying or other intellectual property infringement has occurred. Protective orders should be adopted to prevent harm from disclosure of information pertaining to the privacy and/or security of innocent third parties.

Regulators and professional bodies should promote interoperable rights identifiers and consistent reporting of training content, and facilitate the global exchange of non-confidential compliance information. Where cooperation is implemented, differences in legal rules, e.g., Indian fair dealing, United States fair use, European mining, and United Kingdom copyright exceptions, should be recognized. Legal systems and cooperation mechanisms should work in complement to each other, and establish the facts while recognizing that substantive issues should be determined in accordance with the law governing the claim (Sriram 2025).

Finally, the cost and benefit analysis should be conducted. Duties should be revised in line with new developments in law and technology and protect evidence. Errors should be appealable. Intermediaries should provide scaled duties, and creators should be afforded rights to access reasonably priced dispute resolution mechanisms. Lawmakers should be able to assess whether rights of access are improved without increasing the burden on creators. The improvement should be measurable. For example, the time taken to respond to notices should be improved or made consistent, or the rate at which orders are carried out.

7. Conclusion

The primary focus of the examined cases is concentrated in U.S. case filings and allegations of the use of protected information in the development of models. Overall, the 97 cases do not point to widespread or established cases of unlawful copying, and do not support the likelihood of the success of a legal claim. The 13 selected cases of judicial activity also support the need to evaluate cases, allegations, and merits separately. Not every act of training a large language model is an infringing act. Likewise, not every large language model outputs an expression. Training a large language model for a task does not mean that the large language model can automatically create expressions. Determining the legality of training a large language model is not an absolute science. A more reasonable approach to alleviate the concerns is to implement a combination of acquisitions, reservations, and licensing. This would help settle concerns and provide the flexibility to recognise and defend legitimate business interests. It is also important for researchers to consult with the developers and creators of Large Language Models (LLMs) to gain a more holistic understanding. Evidence and more actionable research should be prioritised. The research is subject to a small sample size.

References

Buick, Adam. 2025. Copyright and AI Training Data—Transparency to the Rescue? *Journal of Intellectual Property Law & Practice* 20: 182–92.

Carlini, Nicholas, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramèr, Borja Balle, Daphne Ippolito, and Eric Wallace. 2023. Extracting Training Data from Diffusion Models. arXiv preprint. https://doi.org/10.48550/arXiv.2301.13188

de la Durantaye, Katharina. 2025. Control and Compensation: A Comparative Analysis of Copyright Exceptions for Training Generative AI. *IIC – International Review of Intellectual Property and Competition Law* 56: 737–70.

Dornis, Tim W., and Nicola Lucchi. 2025. Generative AI and the Scope of EU Copyright Law: A Doctrinal Analysis in Light of C-250/25. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Ginsburg, Jane C. 2025. AI Inputs, Fair Use and the U.S. Copyright Office Report. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Guildea, Brendan. 2026. The Problem of Pleadings in AI Copyright Litigation: Lessons from Getty v Stability (UK) and Doe v GitHub (US). *Journal of Intellectual Property Law & Practice*.

Haidemariam, Tsehaye, and Anne-Britt Gran. 2026. On the Problems of Training Generative AI: Towards a Hybrid Approach Combining Technical and Non-technical Alignment Strategies. *AI & Society* 41: 629–54.

Manteghi, Maryna. 2024. Can Text and Data Mining Exceptions and Synthetic Data Training Mitigate Copyright-Related Concerns in Generative AI? *Law, Innovation and Technology*. https://doi.org/10.1080/17579961.2024.2392928

Mazzi, Francesca. 2024. Authorship in Artificial Intelligence-Generated Works: Exploring Originality in Text Prompts and Artificial Intelligence Outputs through Philosophical Foundations of Copyright and Collage Protection. *The Journal of World Intellectual Property* 27: 410–27.

Pasetti, Marcelo, James William Santos, Nicholas Kluge Corrêa, Nythamar de Oliveira, and Camila Palhares Barbosa. 2025. Technical, Legal, and Ethical Challenges of Generative Artificial Intelligence: An Analysis of the Governance of Training Data and Copyrights. *Discover Artificial Intelligence* 5: 193. https://doi.org/10.1007/s44163-025-00379-6

Peukert, Alexander, and Céline Castets-Renard. 2025. The Copyright Chapter of the EU's Code of Practice for General-Purpose AI Models: A Commentary. *IIC – International Review of Intellectual Property and Competition Law* 56: 1950–93. https://doi.org/10.1007/s40319-025-01664-8

Rai, Priyanshi, Kartik Gupta, Lubna Abdelkhreim Gabralla, and Tanupriya Choudhury. 2025. A Royalty Framework for Copyright Protection and Accountability in AI-Generated Art. *Discover Computing* 28: 252. https://doi.org/10.1007/s10791-025-09766-0

Saw, Cheng Lim, and Bryan Zhi Yang Tan. 2025. Unpacking Copyright Infringement Issues in the GenAI Development Lifecycle and a Peek into the Future. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Sriram, K. S. Amith. 2025. Copyright Facing the Challenges of Generative Artificial Intelligence: Judicial Practice and Legislative Strategies in India, the United States and the European Union. *Journal of Digital Technologies and Law* 3: 598–635.

Tyagi, Kalpana. 2024. Copyright, Text & Data Mining and the Innovation Dimension of Generative AI. *Journal of Intellectual Property Law & Practice* 19 (7): 557–70. https://doi.org/10.1093/jiplp/jpae028

Wu, Hong. 2024. Copyright Protection During the Training Stage of Generative AI: Industry-Oriented U.S. Law, Rights-Oriented EU Law, and Fair Remuneration Rights for Generative AI Training under the UN's International Governance Regime for AI. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Zhang, Victor Jiawei. 2025. Input Out, Output In: Towards Positive-Sum Solutions to AI-Copyright Tensions. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.