Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Rethinking Intellectual Property in the Age of Generative AI: Ethics, Rights and Accountability

Anurag Sharma[1]

sanurag986@gmail.com

K.R. Mangalam University

Research Article

Keywords: generative artificial intelligence, intellectual property, AI ethics, copyright, creator rights, comparative law

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11489398/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Rethinking Intellectual Property in the Age of Generative AI: Ethics, Rights and Accountability

Anurag Sharma[1]*

[1] Assistant Professor, School of Legal Studies, K.R. Mangalam University, Gurugram, Haryana, India

* Corresponding author: Anurag Sharma

Email: sanurag986@gmail.com

ORCID: 0000-0003-1581-3357

Abstract

The rapid commercialisation of generative artificial intelligence (AI) has converted long-standing philosophical debates about authorship, property and fairness into live disputes before courts and legislatures. This article examines the ethical dimensions of AI and intellectual property rights (IPR) through a mixed doctrinal–empirical design. It develops a four-axis analytical framework (Consent, Compensation, Credit and Accountability) and tests it against secondary data drawn exclusively from verifiable public sources: the World Intellectual Property Organization (WIPO) patent landscape on generative AI, the Stanford AI Index 2025, curated litigation trackers, court judgments and settlement orders from the United States, the United Kingdom, Germany and India, creator surveys by the Society of Authors, an economic study commissioned by CISAC, and the foundational algorithmic-bias audit by Buolamwini and Gebru. The findings show that (i) generative-AI patenting is highly concentrated, with China-based inventors accounting for roughly 71 per cent of 54,000 patent families filed in 2014–2023, while India accounts for about 2.5 per cent; (ii) United States copyright suits against AI developers more than doubled, from 46 to 101, between June 2025 and May 2026; (iii) the US$1.5 billion Bartz v. Anthropic settlement, at about US$3,100 per work, equals roughly 2 per cent of the statutory ceiling yet is four times the statutory minimum; (iv) harm to creators is concentrated in readily substitutable segments, with 36 per cent of surveyed translators and 26 per cent of illustrators reporting lost work; and (v) the leading global AI-ethics guidelines, although convergent on transparency (73 of 84 documents), do not treat creator consent or remuneration as a distinct principle. Comparative analysis of seven jurisdictions reveals divergent judicial answers to the same ethical question, most recently illustrated by the Delhi High Court's interim refusal in ANI v. OpenAI (24 July 2026) and the Munich Regional Court's contrary reasoning in GEMA v. OpenAI (11 November 2025). The article concludes by proposing a layered governance model that combines training-data transparency, statutory remuneration with collective management, calibrated human-contribution thresholds for authorship, and shared accountability for outputs, with specific attention to the equity interests of emerging economies.

Keywords: generative artificial intelligence; intellectual property; AI ethics; copyright; creator rights; comparative law

1. Introduction

1.1 Background and problem statement

Intellectual property (IP) law has always been an exercise in applied ethics. Behind each doctrine sits a moral judgment about who deserves to control, profit from or be credited for an intangible creation, and about how far private control should yield to public access. For most of the twentieth century these judgments were stable because the producers of creative and inventive output were assumed to be human beings. Generative AI disturbs that assumption at three points simultaneously. First, systems are trained on enormous corpora of human-made works, usually without the knowledge of the people who made them. Second, the systems themselves generate text, images, music, code and molecular designs that may or may not qualify as protectable subject matter. Third, the outputs can substitute for the very creators whose works made the system possible.

The scale of the phenomenon makes the ethical question practical rather than speculative. The WIPO Patent Landscape Report on Generative AI documented about 54,000 generative-AI patent families published between 2014 and 2023, more than a quarter of them in 2023 alone 1. The Stanford AI Index reported that private investment in generative AI reached US$33.9 billion in 2024 and that 78 per cent of surveyed organisations were using AI in at least one business function 2. At the same time, litigation has multiplied. By 1 May 2026 a public tracker counted 101 copyright suits against AI companies in the United States alone 3, and a single settlement, *Bartz v. Anthropic PBC*, produced a US$1.5 billion fund for authors whose books were copied from pirate repositories 4.

The problem this article addresses is therefore twofold. Descriptively, the law is answering the same ethical questions in different ways in different places, and often provisionally. Normatively, the ethical principles that dominate the AI-governance literature, such as transparency, fairness and accountability, have not been translated into concrete duties toward the creators whose works train the models. Without that translation, IP law risks becoming either an obstacle to socially valuable innovation or an instrument that quietly transfers value from individual creators to a small number of well-capitalised developers.

1.2 Research questions and hypotheses

The article addresses four testable propositions regarding four research questions.

- RQ1. Are AI-IP claim filing rates outpacing the rate at which courts and legislatures are settling the cases' doctrinal issues? (H1: Litigation volume outpaces the number of merits decisions on AI training).

WIPO, 2024.

Maslej, 2025.

Lee, 2026.

Authors Guild, 2026.

- RQ2. Who is feeling the economic and moral impacts of genAI within the creative economy most sharply? (H2:Harm may fall heaviest on niches in which AI output provides a close substitute for human output, translations, illustrations, production music, etc.)
- RQ3. Do courts and legislatures in various jurisdictions emphasize consent, compensation, credit and responsibility in a different way? (H3: Outcomes differ systematically across jurisdictions based on the normative commitments to property and innovation in each.)
- RQ4. Do the major AI-ethics guidelines address the IP-specific interests of creators? (H4: They fail to see creator consent and remuneration as different principles).

1.3 Contribution and structure

The article contributes three things. First, it proposes a simple, easy to use analytical framework, the 4C model, translating abstract AI ethics into questions that can be put to an IP rule or judgment. Second, it compiles and triangulates from legitimate official sources a dataset of quantitative indicators and visualises it in 6 figures and 3 tables each associated to their respective sources. Third, it provides a cross-jurisdictional reading of 7 jurisdictions with the newest Indian updates rarely documented so far. Section 2 is the conceptual framework, section 3 the literature review, section 4 the methodology and its limitations. Section 5 the data, section 6 the cross-jurisdictional reading. Section 7 the discussion of results in light of the hypothesis, section 8 recommendation, section 9 limitations and future research and section 10 the conclusion.

2. Conceptual and theoretical framework

2.1 Key concepts

Artificial intelligence is used here in the operational sense adopted by most regulators: a machine-based system that infers from its inputs how to generate outputs such as predictions, content or decisions. Generative AI, then, means generative models, chiefly large language models and diffusion models, that generate texts, images, sound, video and code, from prompts. Training data, then, means the corpora that such models are fed on to learn statistical patterns, and intellectual property rights in this context means copyright and related rights, patents, trade secrets and trademarks, with most attention currently being given to copyright.

2.2 Moral foundations of intellectual property

Four classic justifications of IP continue to structure the debate. The labour-desert argument, derived from Locke, holds that those who add their effort to unowned or common resources deserve a property interest

in the result. The personality theory, associated with Hegel, treats creative works as extensions of the author's self and therefore as deserving protection against distortion and unauthorised use. The utilitarian or incentive theory defends exclusive rights as a means of encouraging socially valuable creation and disclosure. The distributive or social-planning approach asks what allocation of rights produces a just and culturally vibrant society (Hughes, 1988; Hettinger, 1989).

Each justification points in a different direction when applied to AI. Labour-desert favours creators whose works are ingested without consent, but it also supports claims by those who labour to build, curate and prompt models. Personality theory supports moral-rights objections to style imitation and misattribution, yet it has little to say about a statistical model that has no self to express. Utilitarian reasoning can favour broad access to training data on the ground that machine learning generates large social benefits, or it can favour strong rights on the ground that unremunerated extraction undermines future human creativity. The distributive perspective raises the question of whether the gains from AI will be shared with the creators, and with the countries, whose cultural output supplies the training material.

2.3 From AI ethics principles to IP questions: the 4C framework

The AI-ethics literature has converged on a recognisable set of principles. Floridi and Cowls (2019) synthesised existing statements into five: beneficence, non-maleficence, autonomy, justice and explicability. The UNESCO Recommendation on the Ethics of Artificial Intelligence, adopted by its member states in November 2021, similarly stresses transparency, fairness, accountability and human oversight 5. These principles are valuable but general. To make them usable in an IP setting, this article distils four operational axes.

- **Consent.** Were the holders of rights in training material asked, or given a real opportunity to refuse? This axis links autonomy and justice to the lawfulness of copying.
- **Remuneration**. When a use is allowed without consent, how do we get a share of the resulting value to the people whose works helped create it? This axis connects ideas about distributive justice with remuneration schemes and licensing markets.
- **Credit.** Are human authors, be they the original authors and makers and the humans making the AI work, named, and is the content emanating from the AI clearly labelled? This axis links dignity and explicability to authorship, attribution and labelling guidelines.
- **Accountability.** Who bears responsibility when a system reproduces protected content, imitates identifiable creators or embeds bias? This axis links responsibility and non-maleficence to liability allocation among developers, deployers and users.

UNESCO, 2021.

The 4C framework is deliberately modest. It does not claim to resolve the underlying normative disputes. Its purpose is analytic: each court judgment, statute or policy proposal examined later in this article can be read as taking a position on one or more of the four axes, and divergences between jurisdictions can be explained as different weightings of the same axes.

Table 1. The 4C analytical framework

Axis	Core question	Linked ethical principles	IP rules most implicated	Indicators used in this chapter
Consent	Were rights holders asked, or able to refuse, before their works were used?	Autonomy; justice	TDM exceptions and opt-outs; fair use and fair dealing; acquisition from pirate sources	Litigation counts; dataset provenance (Books3, LAION); coded decisions
Compensation	If use is allowed without consent, is the resulting value shared with creators?	Distributive justice; beneficence	Statutory and collective licensing; settlements; licensing markets	Bartz per-work figure; creator surveys; CISAC projection
Credit	Are human contributors recognised and AI output identified?	Explicability; dignity	Authorship and inventorship rules; moral rights; labelling	Thaler line of cases; disclosure on registration
Accountability	Who answers for infringing outputs, imitation and bias?	Responsibility; non-maleficence	Secondary liability; operator versus user; trade mark; safe harbours	GEMA and Getty decisions; Gender Shades audit; AI incident counts

3. Literature review

3.1 Training data, fair use and text-and-data-mining exceptions

The most heavily debated question is whether copying works to train a model is lawful without permission. In the United States scholars have argued that machine learning is a paradigmatic transformative use because it extracts non-expressive statistical patterns rather than communicating the original expression (Lemley & Casey, 2021; Sag, 2023). Others caution that the argument weakens when models memorise and regurgitate training examples, or when outputs compete in the same market as the inputs (Samuelson, 2023; Henderson et al., 2023). Empirical machine-learning research gives the second concern a factual basis: Carlini et al. (2023) demonstrated that diffusion models can emit near-copies of individual training images, extracting more than a thousand training examples from a widely used model. Lemley (2024) argues more broadly that generative AI strains copyright's assumptions, because the same technical act, copying at scale, can be benign at the input stage and harmful at the output stage.

Outside the US, the statutory limits are as follows: The EU's Directive 2019/790 established a text-and-data-mining (TDM) exception for research organisations and a wider exception for other users with a rights-holder opt-out (Directive (EU) 2019/790, arts. 3-4). Japan's Article 30-4 allows for the use of information analysis that does not intend to enjoy the expression's contents, and Singapore's 2021 reform introduced a computational data analysis exception. With India having no specific or explicit TDM exception in its Copyright Act, 1957, the

Delhi High Court had to deal with the general fair-dealing provision provided under section 52(1)(a) of the Copyright Act, 1957 in the ANI matter.

3.2 Authorship, inventorship and the human-contribution threshold

Authorship doctrine, in most legal systems, presupposes a human author. Gervais (2020) analyses how the notion of a machine as author collides with the Berne Convention's structure, while Abbott (2020) argues, from a policy rather than doctrinal standpoint, that denying protection to AI-generated output may impede innovation and that AI systems should in some contexts be treated as inventors. Courts have so far rejected the latter position. The United States Court of Appeals for the District of Columbia Circuit held in *Thaler v. Perlmutter*, 130 F.4th 1039 (D.C. Cir. 2025), that copyright requires human authorship, and the Supreme Court declined to review the case on 2 March 2026 6. In patent law, the Federal Circuit, the UK Supreme Court and the European Patent Office's Legal Board of Appeal have each held that an inventor must be a natural person (*Thaler v. Vidal*, 2022; *Thaler v. Comptroller-General*, 2023). The unresolved issue is the degree of human contribution that suffices for AI-assisted works, which the US Copyright Office has addressed through registration guidance and its 2025 report on copyrightability 7.

3.3 Bias, representativeness and the ethics of datasets

A parallel literature treats datasets not as legal objects but as ethical ones. Buolamwini and Gebru (2018) showed that commercial gender-classification systems performed far worse on darker-skinned women than on lighter-skinned men, and traced part of the disparity to benchmark datasets composed overwhelmingly of lighter-skinned faces. Bender et al. (2021) argued that very large web-scraped corpora over-represent dominant voices and encode their prejudices. The connection with IP is under-explored. Decisions about which works may lawfully be used, and on what terms, determine what data developers can assemble, and therefore whose language, culture and likeness the resulting systems reflect. A licensing regime that makes data accessible only to well-funded developers could reduce the diversity of datasets, while an overbroad exception could leave source communities with no say in how their material is used.

3.4 Gap in the literature

Scholarship has been prolific in the realm of doctrine and technical ethics, but weaker on systematic empirical measurement connecting the two. It is also often outdated in respect of the 2025-2026 wave of judgments and settlements, with only scattered efforts for comparison with the developments in the US, Europe and China. This article adds to this discourse by combining verifiable quantitative indices into a common scheme of analysis.

Mayer Brown, 2026.

U.S. Copyright Office, 2025.

4. Research methodology

4.1 Design

The research design is mixed doctrinal-empirical based on secondary data. Doctrinal sources include laws, instruments and court rulings; empirical sources include aggregations of descriptive statistics from published reports, trackers and surveys. The doctrinal and empirical strands are combined in the 4C framework: individual quantitative indicators are attached to the axis they seem to relate to most directly, while individual sources of law are coded for the axis or axes on which they take a position. No primary survey, interview, or experimental data were gathered for this article; it does not attempt causal inference.

4.2 Data sources and validation

Sources only primary/authoritative secondary sources, identifiable and verifiable: inter-governmental organisations (WIPO, UNESCO), academia (Stanford HAI), peer-reviewed research (within or post conference only), court judgment and order, government working paper, professional associations, acceptable legal-industry commentary. Each piece of information in the article was checked against at least one original/institutional publication and, if possible, a second source of independent information. The accessing and verification occurred October 2026. Table 2 identifies the main source and – if used – how it is used. In those instances where the source produces a projection and not an observation (the CISAC projection for 2028), the text states 'projection'.

Table 2. Principal data sources and their use

Source	Type	Use in this chapter	Key figures drawn
WIPO Patent Landscape Report on Generative AI (2024)	Intergovernmental report	Concentration of innovation	About 54,000 families; China 38,210; India 1,350
Stanford HAI AI Index Report 2025	Academic annual report	Investment, adoption, incidents	US$109.1bn US investment; 233 incidents in 2024
Lee (2026), litigation tracker	Curated public tracker	H1: litigation growth	46 suits (Jun 2025) to 101 (May 2026)
Judgments and orders: Bartz, Kadrey, Ross, Getty, GEMA, Thaler, ANI	Primary legal sources and legal-industry commentary	H3: coded outcomes	Seven outcomes, Feb 2025 – Jul 2026
Society of Authors (2024); Design Week report on Association of Illustrators survey	Creator surveys	H2: harm to creators	26% / 36% lost work; 32% lost work, average £9,262
CISAC / PMP Strategy (2024)	Commissioned economic study	H2: projected revenue at risk	24% music; 21% audiovisual; €22bn cumulative
Buolamwini & Gebru (2018)	Peer-reviewed audit	Bias and datasets	Error rates up to 34.7%
Jobin, Ienca & Vayena (2019)	Peer-reviewed content analysis	H4: ethics guidelines	84 documents; transparency in 73
DPIIT (2025); Regulation (EU) 2024/1689	Government and legal instruments	Comparative analysis	Hybrid licence proposal; Art. 53 duties

4.3 Coding and analysis

For H1, litigation counts from a longitudinal public tracker were compared across five dated snapshots and set against the number of merits decisions on AI training that the same tracker reported for early 2026. For H2, creator-side indicators from two surveys and one economic projection were compared by segment. For H3, seven judicial or quasi-judicial outcomes between February 2025 and July 2026 were coded by jurisdiction, stage of the AI life cycle (acquisition, training, output or authorship), and the 4C axis on which the decision turned (Table 4). For H4, the frequency of principles in the review of 84 AI-ethics guideline documents by Jobin et al. (2019) was examined for any principle that corresponds to creator consent or remuneration. Percentages and ratios are computed by the author from the source figures and rounded as indicated.

4.4 Limitations of the method

Three limitations deserve emphasis at the outset. First, litigation trackers are curated by individuals or firms and may define a case differently from a court docket. Second, several indicators come from advocacy bodies, such as CISAC and the Society of Authors, whose samples are self-selected and whose projections rest on modelling assumptions. They are used here as indicators of the direction and relative size of harms, not as precise measures. Third, the case sample in Table 4 is small and purposive rather than random; it captures the most prominent decisions rather than all decisions. The findings should accordingly be read as structured evidence supporting or qualifying the hypotheses, not as statistical proof.

5. Empirical findings

This section reports the quantitative and doctrinal evidence in seven parts. Table 3 summarises the principal indicators and the hypothesis each informs; the subsections that follow supply the detail.

Table 3. Summary of empirical indicators and hypothesis assessment

Test	Indicator	Value	Source	Assessment
H1	US copyright suits against AI companies	46 (30 Jun 2025) to 101 (1 May 2026): +120%	Lee (2026)	Supported
H1	Suits per decision on AI training	About 15 (107 suits; 7 decisions)	Lee (2026)	Supported
H2	Creators reporting lost work to generative AI	Translators 36%; illustrators 26%	Society of Authors (2024)	Supported; self-selected sample
H2	Revenue at risk by 2028	Music 24%; audiovisual 21%; dubbing and subtitling 56%	CISAC (2024)	Supported; projection
H3	Coded judicial outcomes (infringement and liability)	3 developer-favouring; 2 rights-holder-favouring; 1 mixed	Table 4	Supported; small sample
H4	Most frequent ethics principle; creator-IP principle	Transparency 73 of 84; no creator-IP principle among eleven	Jobin et al. (2019)	Supported; pre-generative AI
Context	Generative-AI patent families	About 54,000; China about 71%; India about 2.5%	WIPO (2024)	Concentration

Context	Bartz settlement	US$1.5bn; 482,460 works; about US$3,109 per work	Authors Guild (2026); Pearl Cohen (2026)	Price signal

5.1 Concentration of generative-AI innovation

The WIPO patent landscape provides the most authoritative global count of generative-AI inventions. Between 2014 and 2023 it identified about 54,000 patent families, together with more than 75,000 scientific publications; more than 25 per cent of the patent families and over 45 per cent of the publications appeared in 2023 alone, and generative AI still represented only about 6 per cent of all AI patents 8. Filings grew roughly eightfold after 2017, the year the transformer architecture that underlies modern language models was introduced. The geography of this activity is strikingly uneven (Figure 1). China-based inventors account for 38,210 families, roughly 71 per cent of the 54,000 total and about six times the United States figure of 6,276. The Republic of Korea (4,155), Japan (3,409) and India (1,350) follow. Among applicants, Tencent (2,074 families), Ping An Insurance (1,564) and Baidu (1,234) lead, and Chinese entities occupy six of the top ten positions 9.

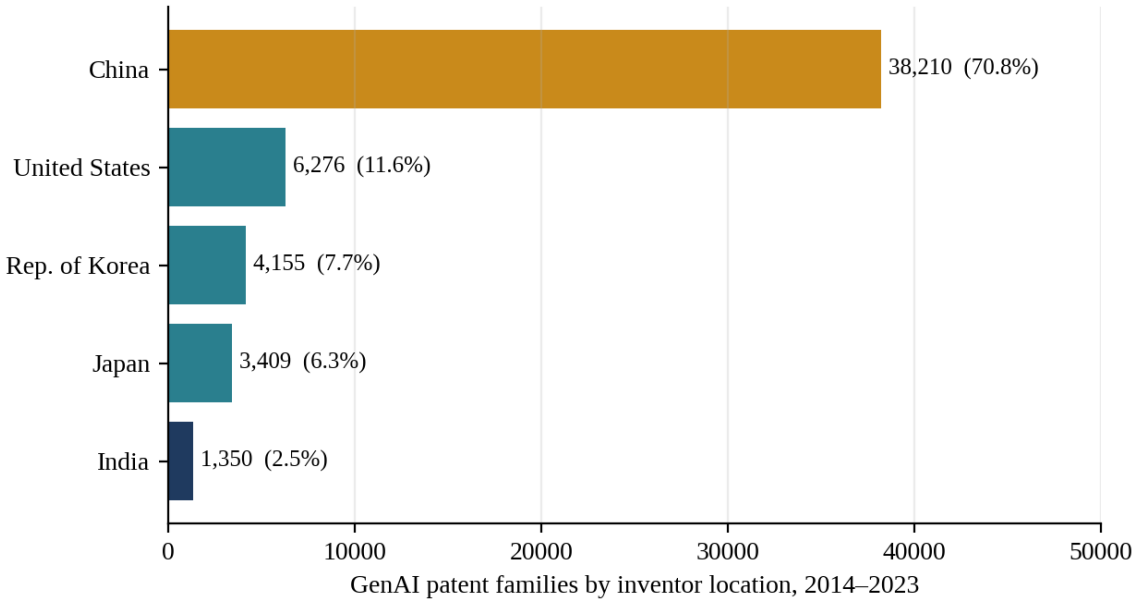


Figure 1. Generative-AI patent families by inventor location, 2014–2023. Source: WIPO (2024). Percentages are the author's calculation against the reported total of about 54,000 families; because a family can have inventors in more than one location, shares are approximate.

Patent activity is only one channel of capacity. The Stanford AI Index shows that private AI investment in the United States reached US$109.1 billion in 2024, nearly twelve times China's US$9.3 billion and twenty-four times the United Kingdom's US$4.5 billion, and that nearly 90 per cent of notable AI models in 2024 originated in industry rather than academia 10. Global private investment in generative AI alone

WIPO, 2024.

WIPO, 2024.

Maslej, 2025.

was US$33.9 billion. Taken together, the two datasets show that inventive output is concentrated in China while investment capital is concentrated in the United States, and that both are far from the countries, including India with about 2.5 per cent of patent families, whose languages and cultural output are being used as training material. For the 4C framework the point is distributive: the entities with the greatest ability to negotiate or contest rules on training data are few and well resourced, whereas rights holders are numerous and dispersed.

5.2 Litigation as an indicator of unresolved ethical conflict

Litigation volume is a visible proxy for the intensity of disagreement. A public tracker maintained by Professor Edward Lee recorded 46 copyright suits against AI companies in the United States on 30 June 2025, 47 on 12 August 2025, 65 on 5 December 2025, 85 on 25 February 2026 and 101 on 1 May 2026 11. The count therefore rose by about 120 per cent in ten months (Figure 2). The same tracker reported 112 suits worldwide by 25 February 2026, and in an earlier post, when the worldwide total stood at 107, it noted that only seven decisions on AI training had so far been issued. That is roughly fifteen suits for each decision, which supports H1: disputes are accumulating far faster than doctrine is being settled.

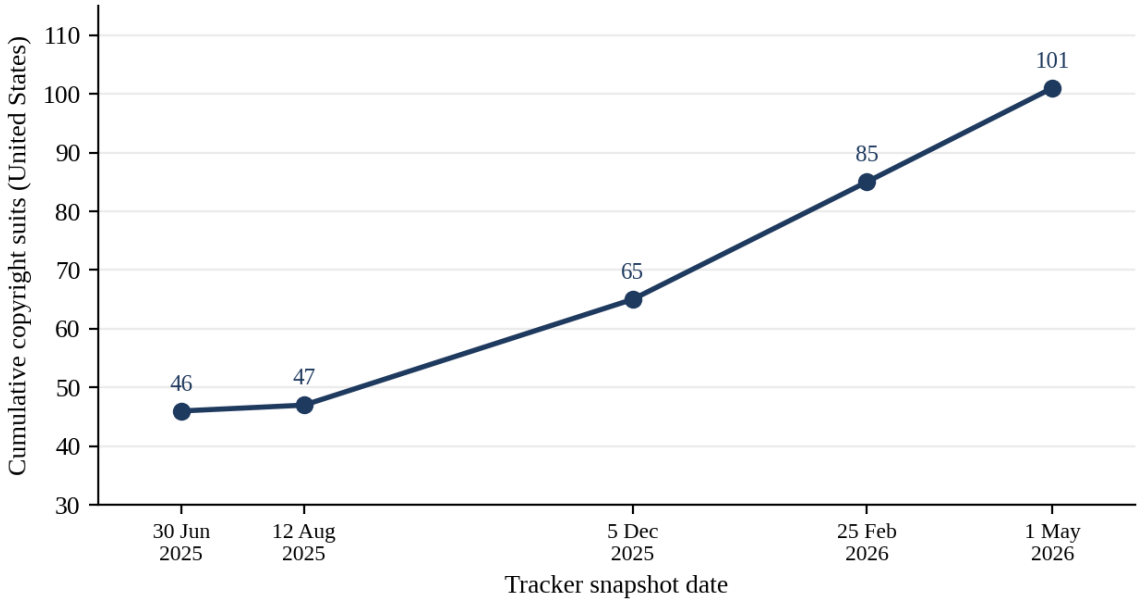


Figure 2. Cumulative copyright suits against AI companies in the United States, June 2025 – May 2026. Source: Lee (2026), compiled by the author from five dated tracker snapshots.

Table 4 codes the seven most prominent judicial outcomes between February 2025 and July 2026. Three observations follow. First, outcomes diverge: of the six decisions that address liability for training, model distribution or outputs, three favoured developers (*Kadrey*, *Getty*, *ANI*), two favoured rights holders (*Ross*, *GEMA*) and one was mixed (*Bartz*). Second, a pattern emerges by stage of the life cycle. The decisions favouring developers concerned training or the distribution of model weights, where a court could characterise the use as non-expressive or the model as not containing a copy. The decisions favouring rights holders concerned a competing product (*Ross*) or the verbatim reproduction of protected content in

Lee, 2026.

outputs (*GEMA*). Third, the decisions rest on different axes: authorship cases turn on Credit, infringement cases on Consent, and output cases on Accountability. No court has yet put Compensation at the centre of a merits ruling; it has been addressed mainly through settlement.

Table 4. Coded judicial and quasi-judicial outcomes, February 2025 – July 2026

Case	Court and date	Stage	Holding in summary	Outcome	Main axis
Thomson Reuters v. Ross Intelligence	D. Del., 11 Feb 2025	Training (non-generative)	Use to build a competing legal-research tool held not fair use	Rights holder	Consent
Bartz v. Anthropic	N.D. Cal., 23 Jun 2025; settlement approved 20 Jul 2026	Acquisition and training	Training on lawfully obtained books held fair use; pirated library copies not; later settled for US$1.5bn	Mixed	Consent; Compensation
Kadrey v. Meta	N.D. Cal., 25 Jun 2025	Training	Fair use for Meta on the record presented	Developer	Consent
Getty Images v. Stability AI	EWHC (Ch), 4 Nov 2025	Model distribution	Secondary copyright claim failed; narrow trade-mark finding; appeal permitted	Developer	Accountability
GEMA v. OpenAI	LG München I, 11 Nov 2025	Output and memorisation	Memorised lyrics held reproduction; operator responsible; no CJEU referral	Rights holder	Accountability; Consent
Thaler v. Perlmutter	D.C. Cir. 2025; cert. denied 2 Mar 2026	Authorship	Human authorship required for copyright registration	Human-authorship rule	Credit
ANI Media v. OpenAI	Delhi HC, 24 Jul 2026 (interim)	Training and output	Interim injunction refused; training prima facie fair dealing; outputs not substantially similar	Developer (interim)	Consent; Compensation

Two cautions apply. The sample is small and purposive, so the pattern by life-cycle stage is a hypothesis for further study rather than a rule. And several outcomes are provisional: the *ANI* ruling is an interim decision on an injunction application and the main suit remains pending 12, Getty has been granted permission to appeal on the secondary-infringement issue 13.

5.3 Training data, consent and compensation

Consent in practice has been difficult to establish because of the way training corpora are assembled. The Books3 collection, built from a pirate library, contained roughly 191,000 books according to an investigation published in *The Atlantic* 14. The LAION-5B dataset compiled 5.85 billion image–text pairs

IP Helpdesk, 2026.

AIPPI, 2026.

Reisner, 2023.

by scraping the web 15. In neither case was the permission of individual rights holders sought. The Hamburg Regional Court held at first instance in *Kneschke v. LAION* (2024) that creating the LAION dataset fell within the German research exception for text and data mining, a result that rests on the scientific-research character of the defendant and does not necessarily extend to commercial model developers.

The *Bartz v. Anthropic* settlement is the clearest available price signal for compensation. The certified class covered 482,460 works downloaded from the pirate libraries LibGen and PiLiMi, and the fund of US$1.5 billion implies about US$3,109 per work before fees (author's calculation), consistent with the court's estimate of about US$3,000 (Pearl Cohen, 2026; Authors Guild, 2026). Claims were filed for 440,490 works, a participation rate of 91.3 per cent that observers described as unusually high for a class settlement 16. Bloomberg Law reported that US$122 million, about 8 per cent of the fund, would go to counsel and litigation costs (Bloomberg Law, 2026). Figure 3 positions the per-work figure against the statutory range under United States law: it is about four times the US$750 statutory minimum, which the court itself noted, yet only about 2 per cent of the US$150,000 maximum for wilful infringement.

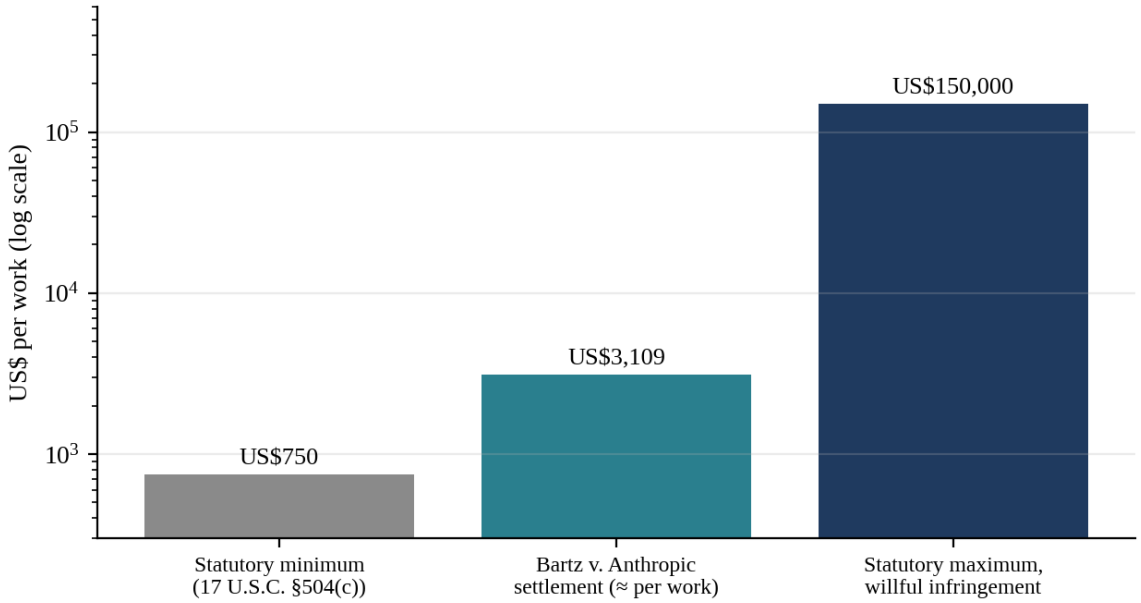


Figure 3. Per-work compensation benchmarks, United States. Statutory range under 17 U.S.C. § 504(c); settlement figure is the author's calculation (US$1.5 billion ÷ 482,460 works). Sources: Authors Guild (2026); Pearl Cohen (2026).

Three ethical points arise. First, the settlement introduced a provenance norm: the dispute turned on how the books were acquired, and the underlying fair-use ruling (*Bartz v. Anthropic PBC*, 2025) distinguished training on lawfully obtained copies from the building of a permanent library from pirated ones. The pirated datasets must be destroyed under the settlement 17. Second, the settlement is retrospective and

---

Schuhmann, 2022.

Writer Beware, 2026.

Pearl Cohen, 2026.

narrow. The final approval order leaves claims about outputs and future conduct outside the release 18, so it compensates for past acquisition without establishing a licence for future training. Third, memorisation research shows that output-stage risk is real rather than hypothetical. Carlini et al. (2023) extracted over a thousand training images from a diffusion model, and the Munich Regional Court in *GEMA v. OpenAI* found that a language model had memorised and could reproduce song lyrics 19. Where a model can regenerate protected content, the argument that training is merely non-expressive learning loses force, and Accountability, not only Consent, becomes the operative axis.

5.4 Harm to creators

H2 predicts that harm will be concentrated where AI output substitutes closely for human work. Three independent data sources support the prediction (Figure 4). In January 2024 the UK Society of Authors surveyed its 12,500 members and received 787 responses. Twenty-six per cent of illustrators and 36 per cent of translators reported having already lost work to generative AI, and 37 per cent of illustrators and 43 per cent of translators said the value of their income had fallen 20. About 94 to 95 per cent of respondents wanted acknowledgment or compensation when their work is used to build generative systems. A separate early-2025 survey by the Association of Illustrators, conducted for the UK government's copyright consultation, found that over 32 per cent of respondents had lost work, averaging £9,262 per affected artist (Design Week, n.d.). The survey data also reveal an element of pressure: 8 per cent of translators and 5 per cent of illustrators said they had used generative AI because a publisher or commissioner asked them to (European Writers' Council, 2024).

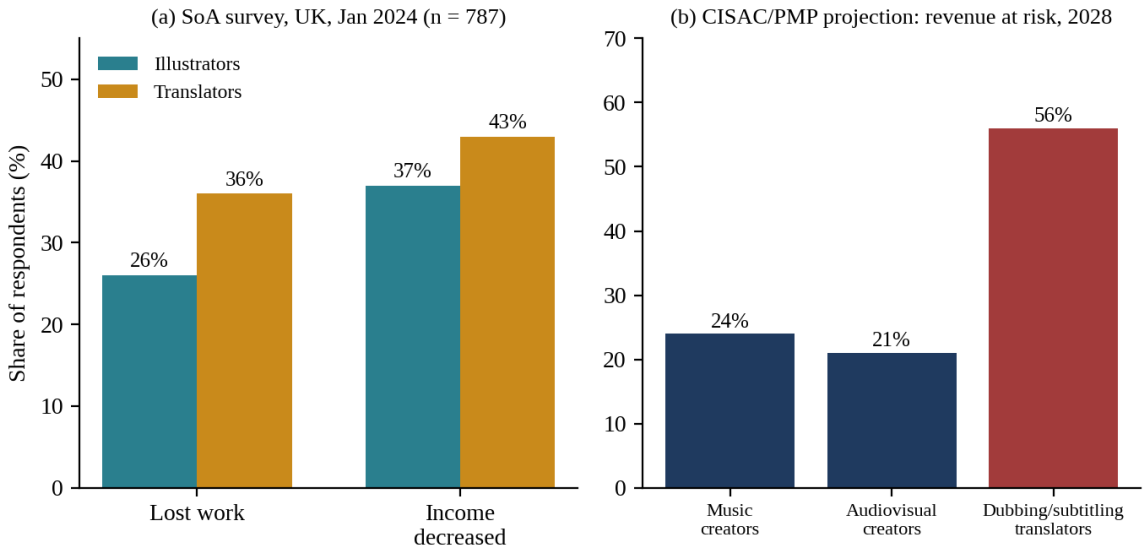


Figure 4. Indicators of creator harm. (a) Society of Authors survey (Jan 2024, n = 787). (b) Projected share of revenue at risk by 2028 in the CISAC/PMP Strategy study. Sources: Society of Authors (2024); CISAC (2024).

Let's Data Science, 2026.

Norton Rose Fulbright, 2025.

Society of Authors, 2024.

The economic projection commissioned by CISAC and carried out by PMP Strategy points the same way. It estimates that by 2028 music creators will have 24 per cent of their revenues at risk and audiovisual creators 21 per cent, a cumulative loss of €22 billion over five years (€10 billion in music, €12 billion in audiovisual), while the market for AI-generated music and audiovisual content grows from about €3 billion to €64 billion. Translators and adaptors working in dubbing and subtitling face the highest exposure at 56 per cent, and screenwriters and directors 15 to 20 per cent (CISAC, 2024; SecurityBrief, 2024). The ordering of these figures supports H2: Exposure is greatest where machine output is most readily interchangeable.

The Evidence has shortcomings: author specific rates for the Society of Authors are around 6 per cent of membership (787/12500 in authors' own calculation) - so the respondents affected may be over-represented, and the CISAC study is a commissioned projection of a body representing rights holders. These caveats do not negate the picture, but suggest that the figures should be used as signs of trend and scale. The ethical import lies in the combination of near-unanimous creator demand for consent, credit or payment with measurable early-stage losses: the Compensation axis is not an abstract concern but a present one.

5.5 Authorship, inventorship and credit

On the Credit axis the law has moved one way. In Thaler v. Perlmutter the D.C. Circuit ruled that work created entirely by an AI system that contains no human element may not be registered,1 and the Supreme Court refused to hear the case on 2 March 2026.2 The decision is not definitive: Dr Thaler had not argued in the case that any human input was creative, and the court therefore declined to specify how much is enough. According to the US Copyright Office, AI-assisted works may be registered where "a human creator has selected, arranged or modified the material in a way that exhibits at least a modicum of creativity", but mere prompts generally are not sufficient to qualify.3 In a lawsuit currently in progress, a user offers up an image that was modified with more than 600 prompts; the court will have to place a limit.

Other jurisdictions have taken different paths. The Beijing Internet Court held in 2023 that an image created with Stable Diffusion was protectable because the user's choices of prompts, parameters and selection reflected sufficient intellectual input 21. The United Kingdom's Copyright, Designs and Patents Act 1988 already contains a provision for computer-generated works, assigning authorship to the person who makes the arrangements necessary for the creation of the work, and India's Copyright Act, 1957 uses similar language, naming as author the person who causes the work to be created. Neither provision was drafted with generative AI in mind, and their application to prompt-driven systems is untested. In patent law, courts and patent offices in the United States, the United Kingdom and Europe have uniformly held that an inventor must be a natural person (*Thaler v. Vidal*, 2022; *Thaler v. Comptroller-General*, 2023).

The ethical stake is twofold. Denying protection to wholly autonomous output protects the incentive and dignity rationale of copyright, which is premised on human creativity, and prevents the creation of a large pool of machine-made works that could crowd out human ones. But a high, undefined threshold could

---

Li v. Liu, 2023.

disadvantage users who make real creative choices, and a low one could allow mass registration of trivial outputs. A disclosure rule that identifies which elements are AI-generated, as the US Copyright Office requires on registration, is a modest but effective way to serve Credit without resolving the threshold question.

5.6 Bias and the ethics of datasets

The ethical dimensions of AI and IP include fairness in system performance because rules on data access shape the composition of datasets. The Gender Shades audit evaluated three commercial gender-classification systems and found error rates of up to 34.7 per cent for darker-skinned women, against 0.0 to 0.8 per cent for lighter-skinned men 22. The gap ranged from about 21 to 34 percentage points across vendors (Figure 5). The authors traced part of the disparity to benchmark datasets that were overwhelmingly composed of lighter-skinned subjects, reporting shares of roughly 80 to 86 per cent in two widely used benchmarks. The Stanford AI Index recorded 233 AI-related incidents in 2024, a record and a 56.4 per cent increase over 2023 23.

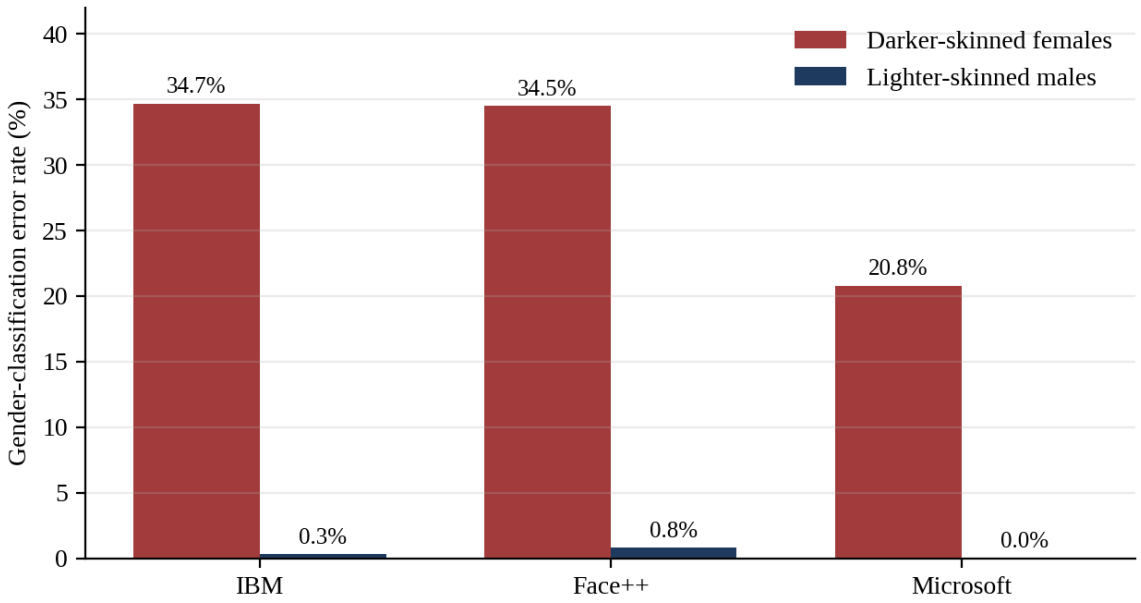


Figure 5. Gender-classification error rates by vendor and subgroup. Source: Buolamwini & Gebru (2018), Gender Shades audit.

The connection with IP is indirect but important. If rights holders can opt out of training use and those who opt out differ systematically from those who do not, the remaining corpus may become less representative; if access is available only through expensive licences, only well-capitalised developers will be able to assemble diverse data. Conversely, an unconditional exception may allow the extraction of cultural material from communities that lack the resources to monitor or contest its use. These are mechanisms rather than demonstrated effects, and empirical work on how opt-out regimes alter dataset composition is still needed. They nevertheless show that a policy choice framed as a copyright question has fairness

Buolamwini, 2018.

Maslej, 2025.

consequences, and that Consent and Compensation rules should be evaluated for their effect on representativeness.

5.7 Do AI-ethics guidelines address creators? Testing H4

Jobin et al. (2019) reviewed 84 AI-ethics guideline documents and identified eleven recurring principles. Transparency appeared in 73 documents (87 per cent), justice and fairness in 68 (81 per cent), non-maleficence and responsibility in 60 each (71 per cent), and privacy in 47 (56 per cent) (Figure 6). None of the eleven principles concerns creators' intellectual property, consent to use of their works, or remuneration. The result supports H4, although with a caveat: the review covers guidelines published up to 2019, before generative AI became a mass-market technology, and later instruments may address the issue more directly.

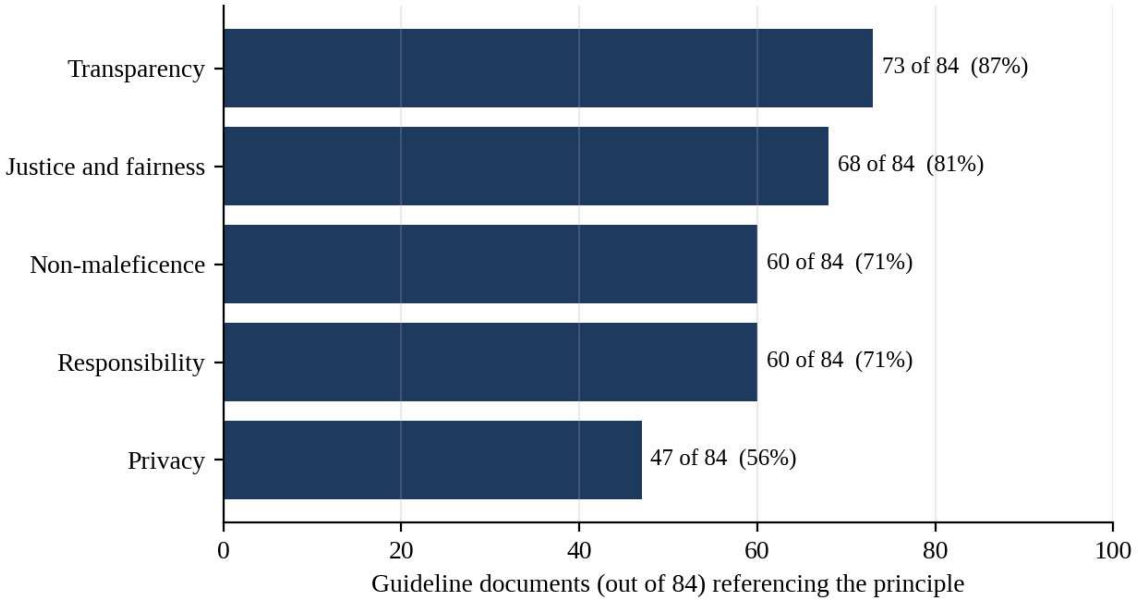


Figure 6. Five most frequent principles across 84 AI-ethics guideline documents. Source: Jobin, Ienca & Vayena (2019). Percentages are the author's calculation.

Hard law has begun to close the gap. Regulation (EU) 2024/1689 obliges providers of general-purpose AI models to maintain a policy for complying with EU copyright law, including respect for rights reservations under the TDM exception, and to publish a sufficiently detailed summary of training content using a template issued by the AI Office. These obligations have applied since 2 August 2025, and the Commission's enforcement powers over general-purpose model providers begin on 2 August 2026 24. The Digital Omnibus, adopted in mid-2026, postpones obligations for Annex III high-risk systems to 2 December 2027 but leaves the general-purpose-model duties unchanged (Orrick, 2026; StudioMeyer, 2026). In effect, the EU has converted the abstract principle of transparency into a creator-facing duty. Whether training summaries will be detailed enough for rights holders to exercise their rights is an empirical question that the first enforcement cycle will answer.

Cloud Security Alliance, 2026.

6. Comparative jurisdictional analysis

Table 5 compares seven jurisdictions on three questions: who is recognised as author, how training use is treated, and what transparency duty applies. The comparison confirms H3: the same ethical question receives different answers, and the differences track each system's baseline commitments.

Table 5. Comparative overview of seven jurisdictions

Jurisdiction	Authorship of AI-assisted output	Training use	Transparency and labelling
United States	Human authorship required; AI elements disclosed on registration; Thaler cert. denied (2026)	Case-by-case fair use; no TDM exception; piracy-sourced copies treated differently (Bartz)	No general federal training-data disclosure statute; disclosure on registration
European Union	Author's own intellectual creation (human) (Infopaq)	TDM exceptions with opt-out (Directive 2019/790, arts. 3–4); GEMA treats memorisation as reproduction	AI Act art. 53: copyright policy and public training summary; art. 50 transparency
United Kingdom	CDPA s. 9(3) computer-generated works; untested for generative AI	TDM exception for non-commercial research only (s. 29A); Getty: weights not an infringing copy	No specific statutory duty; consultation held 2024–25
India	Copyright Act s. 2(d)(vi) (person who causes the work to be created); untested; DPIIT output paper pending	No TDM exception; s. 52(1)(a) fair dealing; ANI interim ruling; DPIIT hybrid licence proposed	No statutory duty yet
China	AI-assisted images protectable where human intellectual input shown (Li v. Liu)	Administrative measures require respect for IP; courts case-by-case	Labelling measures for AI-generated content (2025)
Japan	Human creative contribution required (ACA guidance)	Art. 30-4 information-analysis exception with limits	No statutory duty; official guidance
Singapore	Human authorship presumed; not specifically addressed	Computational data analysis exception (s. 244) with lawful-access condition	No statutory duty

6.1 United States

The United States has no dedicated TDM exception; training use is assessed under the open-ended fair-use doctrine. The district court decisions of 2025 (*Bartz v. Anthropic PBC*, 2025; *Kadrey v. Meta Platforms*, 2025) treated training as highly transformative where the copies were lawfully obtained, but they left room for liability on acquisition from pirate sources and on market-harm theories, including the possibility that outputs dilute the market for human works. The result is case-by-case adjudication, high litigation volume (Section 5.2) and settlement as the main source of compensation. On the 4C axes the system favours innovation-based reasoning on Consent, relies on private ordering for Compensation, and holds firmly to human authorship for Credit.

6.2 European Union and Germany

The EU offers a statutory TDM exception with an opt-out for non-research users, and the AI Act adds the transparency and copyright-policy duties discussed above. EU originality doctrine, which requires the

author's own intellectual creation (*Infopaq*, 2009), already presupposes a human author. The Munich Regional Court in *GEMA v. OpenAI* decided that memorisation of lyrics in model parameters is a reproduction, that output of those lyrics is a communication to the public, and that the operator, not the user, is responsible; it also declined to refer questions to the Court of Justice of the European Union (Norton Rose Fulbright, 2025; D Young & Co, 2025). The EU approach is the most explicit on Accountability and Consent-through-opt-out, though the practical effectiveness of machine-readable opt-outs remains contested.

6.3 United Kingdom

In *Getty Images v. Stability AI* [2025] EWHC 2863 (Ch), the High Court rejected the secondary-infringement claim because the model weights were not themselves a copy of the protected works, and found only narrow trade-mark infringement relating to watermarks (AIPPI, 2026; CMS, 2025). The court did not decide whether training in the UK itself would infringe, because the training took place abroad. The decision shows how territorial and technical features, rather than ethical judgments, can determine outcomes, and it leaves Consent and Compensation for legislative resolution. The government consulted in December 2024 on introducing a TDM exception with a rights-holder opt-out; readers should verify the current legislative position before relying on this article.

6.4 India

India presents two developments that point in different directions. In December 2025 the DPIIT committee, constituted on 28 April 2025, released Part 1 of its Working Paper on Generative AI and Copyright. After reviewing blanket exceptions, TDM exceptions, voluntary licensing and extended collective licensing, it recommended a hybrid model: a mandatory blanket licence permitting AI developers to train on all lawfully accessed works, coupled with a statutory right to remuneration, with royalties payable on commercialisation and collected through a new body, the Copyright Royalties Collective for AI Training (DPIIT, 2025; Mondaq, 2025; Asia IP, 2025). The Ministry of Electronics and Information Technology supported the model, while Nasscom dissented in favour of a TDM exception with opt-out (Storyboard18, 2025). The paper also records the committee's view that unlicensed training may amount to infringement under section 51 of the Copyright Act.

On 24 July 2026, however, Justice Amit Bansal of the Delhi High Court refused ANI's application for an interim injunction against OpenAI. The court upheld its territorial jurisdiction but held, on a prima facie basis, that OpenAI's admitted temporary storage of ANI's works for training fell within the fair-dealing provision in section 52(1)(a), and that the outputs before it were not substantially similar to ANI's content. It also noted that ANI could have blocked OpenAI's crawlers (Intepat, 2026; LiveLaw, 2026; Law.asia, 2026). The decision does not resolve the main suit, which remains pending 25.

Read through the 4C framework, the two developments reveal a tension. The DPIIT proposal places Compensation at the centre and builds a statutory mechanism for it, with Consent replaced by a licence

---

IP Helpdesk, 2026.

that cannot be refused. The *ANI* interim ruling, by treating training as fair dealing, would leave developers free to train without either consent or payment unless Parliament intervenes. The ruling also treats the rights holder's failure to use a technical opt-out as relevant, which, if generalised, would shift the burden of Consent to creators, many of whom lack the means to monitor crawlers. For a country whose creative industries are large and whose AI start-ups depend on affordable data, the ethical case for the DPIIT model is that it avoids both outcomes: it neither blocks training nor leaves creators uncompensated.

6.5 China, Japan and Singapore

China, which dominates generative-AI patenting (Section 5.1), regulates through administrative measures that require providers to respect IP rights and, since 2025, to label AI-generated content (Cyberspace Administration of China et al., 2023, 2025), while its courts have recognised copyright in AI-assisted images with demonstrable human input 26. Japan's Article 30-4 and Singapore's computational data analysis exception reflect an explicit policy choice to prioritise access to data for analysis, with limits where the use unreasonably prejudices the rights holder's interests 27. These permissive models give developers legal certainty but place the Compensation question outside copyright law, a gap that creators' organisations continue to challenge.

7. Discussion

7.1 Assessment of the hypotheses

H1 is proven correct. US lawsuits increased by roughly 120% over 10 months and the tracker's own tally of decisions on AI training rounded to seven when the global total of lawsuits stood at 107. The caveat is that the metrics are inexact: trackers have their own conception of what a case is compared to dockets, and consolidation (for instance the OpenAI multidistrict litigation) may raise or lower the total number of lawsuits without a change in the number of disputes at their core. H2 is largely supported, though with qualification: both survey data and CISAC's projection agree that harm is worst in substitutable segments, but the sample is self-selected and the projection is not based on a representative sample. H3 is supported: the seven mutually exclusive outcomes display contrast, and the contrast has a pattern (training and model distribution seem to help developers; reproduction and competing products seem to help rights-holders), but the sample is too small to consider the pattern law. H4 is supported on the evidence available, with the caveat that the Jobin et al review does not include generative AI, and the EU AI Act was introduced in the months following.

7.2 Evaluating regulatory models against the 4C framework

Table 6 provides a qualitative assessment of the six models against the four axes. The ratings are the author's assessments rather than metrics; their purpose is to be explicit about the trade-offs. They should not be taken as a ranking of the models. Case-by-case fair use, as in the US, maximizes flexibility to the

---

Li v. Liu, 2023.

Agency for Cultural Affairs, 2024.

developer but puts Consent and Compensation in the hands of litigants and settlement negotiations. A broad exception, as in Japan, maximizes legal certainty but performs poorly on Consent and Compensation. An opt-out TDM exception, as in the EU, affords creators a limited form of Consent but imposes a penalty on them and does not afford them Compensation itself. Voluntary licensing protects Consent and Compensation but may be limited by size of developer, and the large incumbents who have the capital to pay for datasets can operate in this space. A consent-first regime provides the strongest protection but makes comprehensive training very difficult and might cause datasets to lose diversity. A statutory blanket licence with remuneration, as proposed in India, performs well on Compensation and access and moderately well on Consent, as it replaces the right to refuse with a right to be paid. Its weaknesses are in the administrative sphere: the collecting agency must be trustworthy, its rates must be reasonable, and it cannot require registration as a condition of payment, which has been flagged as a concern by Indian practitioners.

Table 6. Qualitative assessment of regulatory models for training use (author's analytic judgment)

Model	Example	Consent	Compensation	Developer access	Principal risk
Broad analysis exception	Japan Art. 30-4; Singapore s. 244	Weak	Weak	Strong	Value captured without payment
Case-by-case fair use or fair dealing	United States; India (interim)	Weak to moderate	Partial, via settlement	Strong	Uncertainty and litigation cost
Opt-out TDM exception	EU arts. 3–4	Moderate	Weak	Moderate to strong	Burden on creators; opt-out effectiveness
Voluntary licensing	Publisher and platform deals	Strong	Strong for those licensed	Moderate; weak for small developers	Entrenches incumbents; fragmentation
Consent-first (opt-in)	No major jurisdiction	Strong	Strong	Weak	Narrower, less diverse datasets
Statutory blanket licence with remuneration	DPIIT proposal (India)	Moderate	Strong	Strong	Administration; collecting-body trust; registration barrier

7.3 Equity and the position of emerging economies

The data in Section 5.1 suggest that the ability to build and train models is concentrated, but that their cultural material is global. For developing economies, the central ethical challenge is double extraction: models trained on the works of their creators without remuneration, then sold back to the market. A statutory remuneration scheme limited to works from one jurisdiction or which was based on registration could replicate this asymmetry for authors in languages and cultures with less institutional support. This might help mitigate the risk of unequal asymmetries for creators in languages or fields with weaker institutional backing. India's position is unusual because it is both a large source of training material and a growing

developer; and the DPIIT approach is noteworthy because it explicitly seeks to minimize transaction costs for start-ups while ensuring that creators are paid. Whether it works will depend on whether international rules recognise and coordinate such schemes, since a national licence cannot bind developers who train abroad.

7.4 Reconciling innovation and creator interests

The best argument for permissive rules is that machine learning produces valuable discoveries from an input of patterns not expression. The best argument against is that the benefit is captured by developers whilst the cost is borne by creators whose work gave them the data to train it. The data presented in the current paper indicate both approaches can co-exist: Courts deciding training to be lawful use (Kadrey, ANI) did so while leaving open output; meanwhile, the settlement in Bartz confirms developers will pay significantly for data when provenance is tainted. A governance system that makes training permissible if done for lawful purposes and with lawful access, with transparency and appropriate remuneration, and holds the output generators responsible would be consistent with both positions.

8. Suggestions: moving to a layered model of governance

The results support a layered approach with each layer addressing one or more of the 4C axes. Recommendations are summarised in Table 7; the following paragraphs justify them. Provenance of training data (Consent). For general purpose models, developers should disclose a common machine-readable format-based summary (like that used in the AI Office template under the EU AI Act) of the sources in their training sets. Creators can't exercise a right if they have no knowledge of who might have used their work to train or calibrate the model. This should include the name and provenance of datasets, whether copies were lawfully accessed, and what was done about rights notices.

Compensation For commercial training (Compensation). Jurisdictions that have not opted for a prior consent requirement should accompany this with a statutory equitable remuneration right managed by collective management organisations and contingent on access being lawful. The hybrid model of DPIIT is workable here. Improvements are needed in areas such as: eligibility irrespective of previous registration; published and periodically reviewed royalty rates; minimum distributions per creator to avoid intermediaries claiming all of the fund; and a review clause based on evidence of commercial results. The Bartz figure of about US$3,100 per work, although a negotiated outcome for a specific wrong, offers a reference

point for the order of magnitude that courts consider fair for unauthorised acquisition. Standardised and functional opt-out (Consent). If opt-out is established, the reservation tool must be standards-based, machine-processable and listed in a public registry, and not having a technical opt-out should not be deemed consent for creators who cannot implement one (in particular: individual authors, portfolio authors, and individual artists working with print-based artefacts). ANI's interim logic about banning crawlers demonstrates the potential danger of making technical capacity a requirement for protection. Structured authorship principles and disclosure (Credit). Legislators and registries should uphold the human-authorship principle, resist an sui generis right in output created entirely by autonomous technology, and require disclosure of elements produced by AI in the course of registration and publication. To avoid penalising someone who makes genuine creative choices, guidance should give examples of types of human contribution such as selection, arrangement, adaptation of content, etc.

Mutual obligation to output (Accountability). There should be a duty to exercise reasonable technical means to counter memorisation and regurgitation and liability should follow control: Design and deployed operators will be primarily liable for infringing outputs - as the Munich court found, in the case of a search engine; whereas users will share liability if they explicitly prompt for infringing content. Safe harbours should be contingent on the operator meeting transparency and mitigation obligations. AI generated content should require a label for public facing content.

Equity and global coordination (Compensation, Credit). WIPO and regional organisations should establish minimum standards on training-data transparency and remuneration, and mutual recognition of national collective schemes, to prevent creators in developing countries from being left out of the flow of payments. Funding for capacity building should support smaller nations and developers entering the licensing market. Evidence infrastructure (all axes). Governments and research institutions should support a public, systematically-maintained database of AI-IP disputes, and periodic surveys of creators using probability samples, so that future policy can be based on better data than the advocacy-led indicators we have today.

Table 7. Summary of recommendations

#	Recommendation	4C axis	Lead actor	Horizon
1	Standardised, machine-readable training-data summaries with provenance	Consent	Legislators; regulators	Short term
2	Statutory remuneration for commercial	Compensation	Legislators; collecting	Medium term

	training, with collective administration and no registration precondition		societies	
3	Interoperable opt-out and public reservation registry; no inference of consent from silence	Consent	Regulators; standards bodies	Short term
4	Maintain human-authorship rule; disclosure of AI elements; guidance on contribution	Credit	Copyright offices	Short term
5	Duty to mitigate memorisation; liability follows control; conditional safe harbours; labelling	Accountability	Legislators; courts	Medium term
6	International minimum standards and mutual recognition of national schemes	Compensation; Credit	WIPO; regional bodies	Long term
7	Public litigation database and probability-sample creator surveys	All	Governments; research bodies	Medium term

9. Limitations and directions for future research

This study has limitations that point to research opportunities. It relies on secondary data and does not include primary fieldwork; interviews with creators, developers and collective management organisations in India would illuminate how the proposed hybrid model would work in practice. The litigation sample is small and purposive, and many cases remain undecided, so the conclusions about life-cycle patterns are provisional. Several indicators come from advocacy bodies and may reflect their interests. The legal position is changing quickly: the *Getty* appeal, the main *ANI* suit, the New York Times and other claims consolidated against OpenAI, further tracker updates and the second part of India's DPIIT working paper on outputs may all modify the analysis, and readers should verify current status before relying on any specific ruling. Future research could test the dataset-representativeness mechanism described in Section 5.6 by comparing the demographic and linguistic composition of training corpora before and after opt-out regimes, could develop probability-sample surveys of creator income, and could model the distributional effects of statutory remuneration under alternative royalty rates.

10. Conclusion

The ethical dimensions of AI and intellectual property rights cannot be reduced to a single question of whether training is lawful. They comprise four linked questions: whether creators consented, whether they are compensated, whether human contributors are credited and machine output identified, and who is accountable when systems reproduce or imitate protected work. The evidence assembled in this article shows that these questions are being answered inconsistently and, in many places, provisionally. Generative-AI innovation is concentrated in a few countries, litigation has more than doubled in less than a year, creators in substitutable segments are already reporting lost work and income, and the major ethics guidelines say little about creators' interests. Courts in the United States, the United Kingdom, Germany and India have reached outcomes that favour developers at the training stage and rights holders where outputs reproduce protected content, and settlement rather than adjudication has so far supplied the main form of compensation.

The layered model proposed here, comprising training-data transparency, statutory remuneration with collective administration, standardised opt-outs, calibrated authorship and disclosure rules, shared accountability for outputs, and international coordination, seeks to preserve the social benefits of AI while ensuring that the people whose works make it possible are neither ignored nor left unpaid. India, whose policy-makers are weighing a statutory licence while its courts have signalled a permissive reading of fair dealing at the interim stage, has an opportunity to show that a middle path is possible. The ultimate test of any framework will be whether it sustains both technological progress and the human creativity that supplies its raw material.

Declarations

Conflict of Interest

The author declares that there are no known financial, personal, professional or institutional conflicts of interest that could have influenced the research, analysis, interpretation or conclusions presented in this manuscript.

Funding Declaration

The author declares that this research received no specific financial assistance, grant or funding from any governmental, commercial, institutional or not-for-profit organisation. The research was conducted independently without external financial sponsorship.

Research Ethics and Ethical Approval

The present study adopts a doctrinal, comparative and secondary empirical research methodology. The research incorporates descriptive statistical analysis, including frequencies, percentages, comparative trends, ratios and interpretations of published quantitative findings obtained from publicly accessible and previously published sources.

No primary surveys, interviews, experiments or direct interactions with human participants were undertaken. The statistical findings are derived exclusively from secondary datasets, institutional reports, academic publications, judicial records and other publicly available materials.

Accordingly, informed consent from human participants was not required for this research. The applicability of formal institutional ethics approval is subject to the relevant institutional requirements.

Data Availability Statement

All quantitative and qualitative data examined in this study were obtained from publicly available secondary sources, including reports published by international organisations, academic institutions, regulatory authorities and peer-reviewed publications.

The statistical analysis is descriptive and comparative in nature. No original primary dataset was generated, and no personally identifiable participant information was collected or processed.

The sources supporting the empirical findings are acknowledged in the manuscript and bibliography.

Artificial Intelligence Disclosure

AI-assisted tools disclosure: The author must verify and disclose any use of AI-assisted tools in preparing or revising this manuscript in accordance with the target journal's policy. The author is responsible for the accuracy of all content and citations.

Author Contributions

The author is responsible for the conceptualisation of the research, development of the analytical framework, doctrinal and comparative legal analysis, interpretation of published secondary statistical evidence, formulation of research conclusions and final approval of the manuscript.

Consent for publication: Not applicable; the manuscript does not report identifiable participant information.

Materials availability: Not applicable; no physical research materials were generated.

Code availability: Not applicable; no original software or code is reported.

Corresponding author email and any required declarations must be confirmed before journal submission.

References

Abbott, R. 2020. *The reasonable robot: Artificial intelligence and the law*. Cambridge University Press.

Agency for Cultural Affairs (Japan). 2024. *General understanding on AI and copyright in Japan*. Government of Japan.

AIPPI. 2026, January 23. Getty Images v Stability AI: UK High Court finds no secondary copyright infringement in landmark GenAI ruling. https://www.aippi.org/news/getty-images-v-stability-ai-uk-high-court-finds-no-secondary-copyright-infringement-in-landmark-genai-ruling/

Asia IP. 2025. India proposal on AI developers' use of copyrighted content needs clarity, says lawyer. https://asiaiplaw.com/sector/copyright/india-proposal-on-ai-developers-use-of-copyrighted-content-needs-clarity-says-lawyer

Authors Guild. 2026. Court grants final approval of $1.5 billion Anthropic copyright settlement. https://authorsguild.org/news/court-grants-final-approval-anthropic-copyright-settlement/

Bartz v. Anthropic PBC, No. 4:24-cv-05417 (N.D. Cal. June 23, 2025) (order on fair use); final approval of class settlement (N.D. Cal. July 20, 2026).

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. 2021. On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bloomberg Law. 2026, July 20. Anthropic, authors' $1.5 billion deal earns final approval. https://news.bloomberglaw.com/ip-law/anthropic-authors-1-5-billion-deal-receives-final-approval

Buolamwini, J., & Gebru, T. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research, 81*, 77–91.

Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramèr, F., Balle, B., Ippolito, D., & Wallace, E. 2023. Extracting training data from diffusion models. *32nd USENIX Security Symposium*, 5253–5270.

CISAC. 2024. *Global economic study shows human creators' future at risk from generative AI* [Press release; study by PMP Strategy]. https://www.cisac.org/Newsroom/news-releases/global-economic-study-shows-human-creators-future-risk-generative-ai

Cloud Security Alliance. 2026, May 9. *EU AI Act GPAI: Security compliance before August 2026* [Research note]. https://labs.cloudsecurityalliance.org/research/csa-research-note-eu-ai-act-gpai-enforcement-20260509/

CMS. 2025. *Artificial intelligence and copyright case tracker: Getty Images v Stability AI*. https://cms.law/en/chn/publication/artificial-intelligence-and-copyright-case-tracker/united-kingdom-getty-images-us-inc-ors-v-stability-ai-ltd2

Copyright Act, 1957 (India), ss. 2(d)(vi), 51, 52(1)(a).

Copyright, Designs and Patents Act 1988 (UK), ss. 9(3), 29A.

Cyberspace Administration of China, et al. 2023. *Interim measures for the management of generative artificial intelligence services* (effective Aug. 15, 2023); Cyberspace Administration of China, et al. 2025. *Measures for labeling AI-generated synthetic content* (effective Sept. 1, 2025).

D Young & Co. 2025. The future of AI and copyright law: GEMA v OpenAI and Getty Images v Stability AI. https://www.dyoung.com/de/wissensbank/artikel/ai-copyright-gema-openai-gettyimages-stability

Design Week. (n.d.). How the illustration industry is grappling with AI: A special report. https://www.designweek.co.uk/how-the-illustration-industry-is-grappling-with-ai-a-special-report/

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. *Official Journal of the European Union*, L 130, 92–125.

DPIIT. 2025, December 8. *Working paper on generative artificial intelligence and copyright (Part 1)*. Department for Promotion of Industry and Internal Trade, Government of India.

Eastern Book Co. v. D.B. Modak, (2008) 1 SCC 1 (India).

European Writers' Council. 2024. SoA survey UK: A third of translators and quarter of illustrators losing work to AI. https://europeanwriterscouncil.eu/soa-survey-uk-ai-2024/

Floridi, L., & Cowls, J. 2019. A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

GEMA v. OpenAI, Landgericht München I, Case No. 42 O 14139/24 (Nov. 11, 2025).

Gervais, D. J. 2020. The machine as author. *Iowa Law Review, 105*(5), 2053–2106.

Getty Images (US) Inc. v. Stability AI Ltd. [2025] EWHC 2863 (Ch).

Henderson, P., Li, X., Jurafsky, D., Hashimoto, T., Lemley, M. A., & Liang, P. 2023. Foundation models and fair use. *Journal of Machine Learning Research, 24*(400), 1–79.

Hettinger, E. C. 1989. Justifying intellectual property. *Philosophy & Public Affairs, 18*(1), 31–52.

Hughes, J. 1988. The philosophy of intellectual property. *Georgetown Law Journal, 77*(2), 287–366.

Infopaq International A/S v. Danske Dagblades Forening, Case C-5/08, EU:C:2009:465.

Intepat. 2026. ANI vs OpenAI judgment: What the Delhi High Court actually held. https://www.intepat.com/blog/ani-vs-openai-judgment-delhi-high-court

IP Helpdesk, European Commission. 2026, July 28. Delhi High Court rules on ANI's copyright claims against OpenAI. https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/delhi-high-court-rules-anis-copyright-claims-against-openai-2026-07-28_en

Jobin, A., Ienca, M., & Vayena, E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kadrey v. Meta Platforms, Inc., No. 3:23-cv-03417-VC (N.D. Cal. June 25, 2025).

Kneschke v. LAION e.V., Landgericht Hamburg, 310 O 227/23 (Sept. 27, 2024).

Law.asia. 2026, July 29. OpenAI secures first win in India copyright row against ANI. https://law.asia/openai-copyright-victory/

Lee, E. 2026. *Map of copyright suits v. AI* [Litigation tracker]. ChatGPT Is Eating the World. https://chatgptiseatingtheworld.com/category/map-of-ai-copyright-lawsuits

Lemley, M. A. 2024. How generative AI turns copyright law upside down. *Science and Technology Law Review, 25*(2), 21–44.

Lemley, M. A., & Casey, B. 2021. Fair learning. *Texas Law Review, 99*(4), 743–785.

LiveLaw. 2026, July 24. Delhi High Court rejects ANI's interim plea against ChatGPT, says it would harm public interest. 2026 LiveLaw (Del) 687. https://www.livelaw.in/high-court/delhi-high-court/delhi-high-court-rejects-ani-copyright-suit-open-ai-chatgpt-542736

Li v. Liu, Beijing Internet Court, (2023) Jing 0491 Min Chu No. 11279.

Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., et al. 2025. *The AI Index 2025 annual report*. Stanford Institute for Human-Centered AI. https://hai.stanford.edu/news/ai-index-2025-state-of-ai-in-10-charts

Mayer Brown. 2026, March. Supreme Court denies review in AI authorship case. https://www.mayerbrown.com/en/insights/publications/2026/03/supreme-court-denies-review-in-ai-authorship-case

Mondaq. 2025. Exploring the DPIIT's working paper on generative AI and copyright. https://www.mondaq.com/india/copyright/1718746/exploring-the-dpiits-working-paper-on-generative-ai-and-copyright

Norton Rose Fulbright. 2025. Germany delivers landmark copyright ruling against OpenAI: What it means for AI and IP. https://www.nortonrosefulbright.com/en-cn/knowledge/publications/656613b2/germany-delivers-landmark-copyright-ruling-against-openai-what-it-means-for-ai-and-ip

Orrick. 2026, May. The EU's Digital Omnibus on AI: 7 key changes you need to know. https://www.orrick.com/en/Insights/2026/05/EUs-Digital-Omnibus-on-AI-7-Key-Changes-You-Need-to-Know

Pearl Cohen. 2026. Federal court approves $1.5 billion Anthropic copyright settlement: Largest in history. https://www.pearlcohen.com/federal-court-approves-1-5-billion-anthropic-copyright-settlement-largest-in-history

Let's Data Science. 2026. Judge approves Anthropic's $1.5 billion copyright settlement. https://letsdatascience.com/news/judge-approves-anthropics-15-billion-copyright-settlement-dc3eee91

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L, 2024/1689.

Reisner, A. 2023, September 25. Revealed: The authors whose pirated books are powering generative AI. *The Atlantic*.

Sag, M. 2023. Copyright safety for generative AI. *Houston Law Review, 61*(2), 295–347.

Samuelson, P. 2023. Generative AI meets copyright. *Science, 381*(6654), 158–161. https://doi.org/10.1126/science.adi0656

Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., et al. 2022. LAION-5B: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems 35 (Datasets and Benchmarks Track)*.

SecurityBrief New Zealand. 2024. Generative AI puts AUD$35 billion of creators' income at risk. https://securitybrief.co.nz/story/generative-ai-puts-aud-35-billion-of-creators-income-at-risk

Society of Authors. 2024, April 11. SoA survey reveals a third of translators and quarter of illustrators losing work to AI. https://www2.societyofauthors.org/2024/04/11/soa-survey-reveals-a-third-of-translators-and-quarter-of-illustrators-losing-work-to-ai/

Storyboard18. 2025, December. DPIIT releases working paper on copyright and AI, proposes hybrid licensing model. https://www.storyboard18.com/digital/dpiit-releases-working-paper-on-copyright-and-ai-proposes-hybrid-licensing-model-85551.htm

StudioMeyer. 2026. The EU AI Act in 2026: Reading the law after the Omnibus. https://studiomeyer.io/en/blog/eu-ai-act-2026-after-the-omnibus

Thaler v. Comptroller-General of Patents, Designs and Trade Marks [2023] UKSC 49.

Thaler v. Perlmutter, 130 F.4th 1039 (D.C. Cir. 2025), cert. denied, No. 25-449 (U.S. Mar. 2, 2026).

Thaler v. Vidal, 43 F.4th 1207 (Fed. Cir. 2022).

Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., No. 1:20-cv-613 (D. Del. Feb. 11, 2025).

UNESCO. 2021. *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization.

U.S. Copyright Office. 2025. *Copyright and artificial intelligence, Part 2: Copyrightability*. Library of Congress.

WIPO. 2024. *WIPO patent landscape report: Generative artificial intelligence (GenAI)*. World Intellectual Property Organization. https://www.wipo.int/pressroom/en/articles/2024/article_0009.html

Writer Beware. 2026, July 23. Anthropic settlement update: Final settlement approved. https://writerbeware.blog/2026/07/23/anthropic-settlement-update-final-settlement-approved/