
GENERATIVE AI TRAINING AND COPYRIGHT INFRINGEMENT: A COMPARATIVE ANALYSIS OF THE INDIA, UK, US AND EU

Ankit Shaw, Indian Institute of Technology Kharagpur, India

ABSTRACT

Generative artificial intelligence systems are trained on corpora assembled largely through the automated collection of text, images and other expressive material from the internet, a substantial proportion of which is protected by copyright. This has generated an unresolved and rapidly evolving controversy: does the unlicensed use of copyrighted works to train a generative model infringe copyright, and does any statutory exception excuse it? This paper undertakes a comparative doctrinal analysis of how four major jurisdictions the United States, the European Union, the United Kingdom and India are answering that question. It examines the United States' fact-intensive fair use jurisprudence, including the 2025 split rulings in *Bartz v. Anthropic* and *Kadrey v. Meta* and the pending appellate reconsideration of *Thomson Reuters v. Ross Intelligence*; the European Union's opt-out text-and-data-mining regime under the Digital Single Market Directive as layered with the transparency obligations of the AI Act; the United Kingdom's narrower research-only exception and its unresolved, repeatedly deferred reform process; and India's closed-list fair-dealing framework, examined afresh through the Delhi High Court's July 2026 interim order in *ANI Media v. OpenAI*. The analysis finds that despite sharply divergent statutory starting points, the four jurisdictions are converging functionally on three regulatory levers: purpose-based transformativeness qualified by market-substitution harm, the lawfulness of the underlying source copy, and mandatory transparency about training data. The paper concludes with jurisdiction-specific recommendations, with particular attention to the policy choices confronting India's ongoing Copyright Act review.

Keywords: Generative artificial intelligence, copyright infringement, fair use, fair dealing, text and data mining, AI training data, comparative copyright law.

I. Introduction

Since the public release of large-scale generative systems such as GPT-3 and its successors, generative artificial intelligence (AI) has moved from a research curiosity to a general-purpose technology embedded in search engines, office software and consumer applications used by hundreds of millions of people. These systems large language models that generate text, and diffusion models that generate images, audio and video share a common technical precondition: they must first be trained on a very large corpora of existing human-created material in order to learn the statistical structure of language, image or sound. In practice, developers have assembled these corpora chiefly by crawling and scraping publicly accessible material from the open internet, supplemented by licensed datasets, digitised books and, in at least one well-documented instance, mirrors of pirated book repositories.¹ A substantial share of that scraped and downloaded material is protected by copyright, and very little of it was licensed for the purpose of training a machine-learning model at the time it was collected.

This mismatch between how generative AI is built and how copyright law was designed to operate has produced one of the most consequential and unsettled bodies of litigation and regulatory activity in contemporary intellectual property law. Copyright statutes across the world grant authors an exclusive right to reproduce their works and to prepare derivative works based upon them; training a generative model requires, at a minimum, that the work be reproduced copied into a training corpus and processed, and arguably that derivative statistical representations of it be created. Whether that reproduction is excused, because it falls within an exception such as fair use or fair dealing, because it constitutes a lawful form of text and data mining, or because no legally cognisable “copy” of the work survives inside a trained model, is a question to which national laws, courts and regulators are giving markedly different answers.

This paper undertakes a comparative doctrinal analysis of that question across four jurisdictions chosen for their contrasting legal architectures and practical significance to the global AI industry: the United States, home to most leading AI developers and the largest volume of AI-copyright litigation; the European Union, which has adopted the most comprehensive dedicated regulatory framework in the AI Act; the United Kingdom, a major creative-industries economy

¹ *Bartz v. Anthropic PBC**, No. 3:24-cv-05417 (N.D. Cal. 23 June 2025) (order on cross-motions for summary judgment).

that has oscillated between competing policy options without yet legislating; and India, a rapidly growing AI market whose closed-list fair-dealing tradition has just received its first significant judicial gloss in the Delhi High Court's interim order in *ANI Media Pvt. Ltd. v. OpenAI Inc.*, delivered on 24 July 2026.²

The paper proceeds as follows. Section II explains the technical process of generative AI training and identifies the copyright interests it implicates, distinguishing the training (input) stage from the generation (output) stage. Section III sets out the comparative methodology used. Sections IV through VII examine, in turn, the legal position in the United States, the European Union, the United Kingdom and India, covering statutory exceptions, leading case law, and current regulatory activity. Section VIII draws these accounts together into a structured comparison. Section IX considers the practical challenges this fragmented landscape creates and offers policy recommendations, with particular attention to India's ongoing review of its Copyright Act, 1957. Section X concludes.

Two caveats define the paper's scope. First, the analysis concentrates on the training stage whether copying works into a training dataset and using them to adjust model parameters is itself infringing rather than on the comparatively less litigated question of whether a specific output of a generative system infringes a specific work it resembles; where output-stage doctrine bears directly on the training-stage analysis, as it increasingly does through the "market dilution" theory discussed in Section IV, it is addressed accordingly. Second, given the pace of change in this field, the law is stated as of July 2026; several of the proceedings discussed below, including the Third Circuit's pending decision in *Thomson Reuters v. Ross Intelligence* and the substantive trial in *ANI Media v. OpenAI*, remain undecided at the time of writing.

II. Generative AI Training: The Technical and Legal Interface

A generative model is typically trained by assembling a corpus for a large language model, often hundreds of billions to trillions of words; for an image model, hundreds of millions to billions of image-text pairs and repeatedly presenting fragments of that corpus to a neural network, adjusting millions or billions of numerical parameters ("weights") so the network becomes progressively better at predicting the next element in a sequence. The United States

² *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr.**, Delhi High Court, interim order dated 24 July 2026 (Bansal, J.).

Copyright Office's third report on the subject describes this process as implicating the reproduction right at several distinct points: when a work is first downloaded and stored, when it is copied into curated "data mixes" for a specific training run, and, more contestably, in the transient copies made as the work is processed during training itself.³ Developers and their critics disagree sharply about what, if anything, survives inside a trained model. Developers characterise the resulting weights as compressed statistical abstractions rather than a retrievable copy of any given work, while rightsholders point to documented instances of models reproducing lengthy passages of training text under certain prompts as evidence that some works are, in a meaningful sense, retained. Courts examining this question, discussed further in Section VI, have so far tended to accept the "no stored copy" characterisation of the trained model itself while treating the antecedent training corpus as a conventional, actionable copy.⁴

Copyright's engagement with this process operates at two analytically distinct stages, which this paper keeps carefully separate. The input stage concerns whether assembling and using the training corpus itself infringes the reproduction, and in some formulations the adaptation, right in the works it contains, irrespective of what the resulting model later produces. The output stage concerns whether a specific generated output is substantially similar to a specific pre-existing work a conventional infringement inquiry, complicated in the AI context chiefly by scale and opacity rather than by any novel legal principle. Much of the doctrinal innovation examined in this paper fair use's transformative-purpose analysis, the European Union's text-and-data-mining exception, India's fair-dealing categories is addressed to the input stage. Output-stage liability is comparatively settled in principle but practically difficult to establish, since a claimant must show that a particular output resembles a particular protected work closely enough to support a finding of copying, a threshold generalist outputs will often not meet even where the underlying training was extensive and unlicensed.

A further complication specific to the current generation of AI systems is the growing use of retrieval-augmented generation and live web browsing, which allow a model to retrieve and summarise current source material at the time of a user query rather than relying solely on parameters fixed during training. This blurs the input/output distinction further, since an output produced this way draws on a work that may never have formed part of the training corpus at

³ U.S. Copyright Office, *Copyright and Artificial Intelligence*, Part 3: Generative AI Training (pre-publication version, 9 May 2025).

⁴ Mayer Brown, "Getty Images v Stability AI: What the High Court's Decision Means for Rights-Holders and AI Developers" (20 Nov. 2025).

all; the Delhi High Court's interim order in *ANI Media v. OpenAI*, discussed in Section VII, turned in part on precisely this distinction.⁵

III. Methodology

This paper adopts a doctrinal and comparative legal research methodology. Primary sources legislation, regulations and judicial decisions are analysed directly; where a decision remains under appeal or a regulatory process is incomplete, that procedural posture is noted explicitly rather than treated as settled law. Secondary sources are drawn from official regulatory and governmental reports, including the three-part United States Copyright Office study on *Copyright and Artificial Intelligence*, the United Kingdom's 2026 Report on *Copyright and Artificial Intelligence*, and European Commission guidance on the AI Act, together with contemporaneous professional legal commentary used to verify procedural detail not yet reflected in an official law report.

The four jurisdictions were selected purposively. The United States hosts the largest number of leading generative AI developers and the largest volume of substantively adjudicated training-data litigation, making it the most doctrinally developed body of case law available for study. The European Union is the only one of the four to have adopted a comprehensive, AI-specific statutory instrument alongside a pre-existing harmonised text-and-data-mining exception, offering a contrasting *ex ante* regulatory model to the United States' *ex post*, litigation-driven approach. The United Kingdom, despite sharing a common-law fair-dealing heritage with India, has pursued a materially different and still-unresolved policy path since leaving the European Union, making it a useful intermediate case. India is included both for its scale as one of the world's fastest-growing markets for AI deployment and because its Copyright Act, 1957 represents the closed-list "fair dealing" model as distinct from the open-ended United States "fair use" standard shared by much of the Commonwealth, and because it has, in the course of this paper's preparation, produced its own first judicial ruling on the question.

The comparative analysis in Section VIII is organised around five axes: the structure of the applicable statutory exception; the current state of case law; transparency and disclosure obligations; the treatment of territoriality and jurisdiction; and the direction of ongoing policy

⁵ *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr.**, Delhi High Court, interim order dated 24 July 2026 (Bansal, J.).

reform. The paper does not undertake an empirical or economic analysis of licensing markets, which lies beyond its doctrinal scope, though such markets are referenced where they inform the legal analysis.

IV. The United States

A. Statutory Framework

United States copyright law contains no exception specific to text and data mining or AI training. Any unauthorised reproduction of a copyrighted work must instead be justified, if at all, under the general fair use defence codified at 17 U.S.C. § 107, which directs courts to weigh four non-exclusive factors: the purpose and character of the use, including whether it is commercial and whether it is “transformative”; the nature of the copyrighted work; the amount and substantiality of the portion used; and the effect of the use on the potential market for the original.⁶ Because fair use is an equitable, fact-intensive, case-by-case defence rather than a bright-line exception, the same underlying conduct training a generative model on copyrighted text has produced sharply different outcomes depending on how a given plaintiff’s evidence and a given model’s use case were framed.

The doctrinal foundation for treating AI training as potentially transformative rests on *Authors Guild, Inc. v. Google, Inc.*, in which the Second Circuit held that Google’s mass digitisation of millions of books to power a full-text search-and-snippet tool was fair use because the purpose of the copying, indexing and searching, differed fundamentally from the works’ original expressive purpose.⁷ That reasoning has proved directly transferable to AI training, since a model trained on a book is not, developers argue, a substitute for reading the book but a tool for a functionally different purpose. The Supreme Court’s 2023 decision in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, however, narrowed the availability of a transformative-use finding where a secondary use serves substantially the same commercial purpose as the original and competes with it in the same market, a limitation several of the cases below have had to confront directly.⁸

⁶ U.S. Copyright Office, *Copyright and Artificial Intelligence*, Part 3: Generative AI Training (pre-publication version, 9 May 2025).

⁷ *Authors Guild, Inc. v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015).

⁸ *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023).

B. The 2025 Split Rulings: *Bartz v. Anthropic* and *Kadrey v. Meta*

The first substantive judicial rulings applying fair use squarely to large language model training were handed down within days of each other in June 2025, in the Northern District of California. In *Bartz v. Anthropic*, Judge William Alsup granted summary judgment to Anthropic on two of the three uses at issue: training its models on books, found to be a highly transformative fair use because the model was not reproducing the books but learning generalisable patterns from them, and Anthropic's digitisation of print books it had lawfully purchased, since the resulting digital copy simply replaced a discarded physical original the company already owned.⁹ The court denied summary judgment, however, on Anthropic's retention and use of copies downloaded from pirate repositories, holding that the illegality of how a copy was first obtained is analysed independently and is not cured by a later transformative use.¹⁰ That distinction proved decisive: facing statutory damages exposure across roughly 500,000 pirated works, Anthropic agreed to a US\$1.5 billion class settlement in September 2025, reported to be the largest copyright recovery in United States history, which received final court approval in mid-2026.¹¹ The fair use ruling on lawfully acquired training material was not disturbed by the settlement and stands as the first reasoned judicial precedent on the underlying question.

Two days after Alsup's ruling, in *Kadrey v. Meta Platforms, Inc.*, Judge Vince Chhabria likewise granted summary judgment for Meta on the plaintiffs' claim that training its models on their books was infringing, but on considerably narrower grounds, and used his opinion to signal disagreement with parts of Alsup's reasoning.¹² Chhabria's decision turned specifically on the plaintiffs' failure to develop a record on the fourth fair use factor, market harm, and he indicated that a plaintiff able to show that a model's outputs diluted the market for their specific works, by generating a flood of competing, stylistically similar content, might well defeat a

⁹ *Bartz v. Anthropic* PBC*, No. 3:24-cv-05417 (N.D. Cal. 23 June 2025) (order on cross-motions for summary judgment).

¹⁰ *Bartz v. Anthropic* PBC*, No. 3:24-cv-05417 (N.D. Cal. 23 June 2025) (order on cross-motions for summary judgment).

¹¹ Kluwer Copyright Blog, "The *Bartz v. Anthropic* Settlement: Understanding America's Largest Copyright Settlement" (2026); IBTimes UK, "Anthropic's \$1.5 Billion Book Piracy Settlement Gets Final Approval" (2026).

¹² *Kadrey v. Meta Platforms, Inc.**, No. 3:23-cv-03417 (N.D. Cal. 25 June 2025) (order on motion for summary judgment).

fair use defence even where the training itself was transformative in Alsup's sense.¹³ This "market dilution" theory, distinct from the narrower question of whether an output substantially reproduces a specific work, has since been identified as the most doctrinally significant open question left after the 2025 rulings.¹⁴

C. Thomson Reuters v. Ross Intelligence: A Contrary Result, Now on Appeal

Unlike Bartz and Kadrey, which concerned general-purpose generative chatbots, Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc. concerned an AI-powered legal research tool trained in part on headnotes and the Key Number System from the Westlaw database, built by a direct competitor seeking to displace Westlaw in the market for legal research.¹⁵ In February 2025, Judge Stephanos Bibas, reversing his own earlier tentative view, granted summary judgment to Thomson Reuters, holding that the headnotes were sufficiently original to be protected and that Ross's use of them to train a competing commercial research product was not transformative, since the ultimate purpose of both products was to help lawyers find case law.¹⁶ This was the first American ruling to reject an AI-training fair use defence outright, and its result, turning heavily on direct market substitution between the plaintiff's and defendant's products, is often read alongside Kadrey's market-dilution dictum as marking the outer boundary of the more permissive Bartz approach. The case was certified for interlocutory appeal, and the Third Circuit, sitting as the first federal appellate court to review an AI-training fair use question, heard oral argument on 11 June 2026; its decision, expected to supply the first binding appellate guidance on the issue, remained pending at the time of writing.¹⁷

D. Ongoing Litigation and the Copyright Office's Intervention

Several other major disputes remain unresolved. The New York Times Company's suit against OpenAI and Microsoft, filed in December 2023 and alleging that ChatGPT was trained on and capable of reproducing substantial portions of Times journalism, remains in discovery in the

¹³ *Kadrey v. Meta Platforms, Inc.**, No. 3:23-cv-03417 (N.D. Cal. 25 June 2025) (order on motion for summary judgment).

¹⁴ U.S. Copyright Office, *Copyright and Artificial Intelligence*, Part 3: Generative AI Training (pre-publication version, 9 May 2025).

¹⁵ *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, No. 1:20-cv-00613 (D. Del. 11 Feb. 2025) (order on renewed motions for summary judgment).

¹⁶ *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, No. 1:20-cv-00613 (D. Del. 11 Feb. 2025) (order on renewed motions for summary judgment).

¹⁷ *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, No. 25-2153 (3d Cir.) (oral argument heard 11 June 2026; decision pending).

Southern District of New York more than two and a half years after filing, having generated extensive collateral litigation over preservation of user chat logs and, as of July 2026, a pending motion for sanctions alleging withheld evidence.¹⁸ *Andersen v. Stability AI*, a putative class action brought by visual artists against several AI-image developers, has likewise proceeded for more than three years without a ruling on the ultimate fair use question, though the presiding judge has allowed direct infringement claims to advance to a third amended complaint.¹⁹ Getty Images' parallel infringement claims against Stability AI, originally filed in Delaware and refiled in the Northern District of California in August 2025 after Getty's related but distinct action in the English High Court proceeded to judgment, also remain pending.²⁰

Against this unsettled case law, the Copyright Office issued the third and final part of its *Copyright and Artificial Intelligence* study in May 2025, concluding that whether training on copyrighted works is fair use depends heavily on case-specific facts, while taking the position that large-scale commercial training on works obtained through illegal access, or that produces outputs substituting for those works in their existing markets, exceeds the bounds fair use has traditionally been understood to protect.²¹ The report's release coincided with the controversial removal of the sitting Register of Copyrights, which several commentators linked to the report's guidance, underscoring how politically contested the issue has become at the level of federal policy.²²

E. Legislative Activity

Congress has not resolved the underlying substantive question but has advanced narrower proposals. The bipartisan TRAIN Act, reintroduced in January 2026, would create an administrative subpoena mechanism allowing a rightsholder with a good-faith belief that their work was used in training to compel disclosure of a developer's training records a

¹⁸ *The New York Times Co. v. Microsoft Corp. & OpenAI Inc.*, No. 1:23-cv-11195 (S.D.N.Y., filed 27 Dec. 2023); Al Jazeera, "NYT-led group asks court to sanction OpenAI in US copyright dispute" (9 July 2026).

¹⁹ *Andersen v. Stability AI Ltd.*, No. 3:23-cv-00201 (N.D. Cal., filed 13 Jan. 2023).

²⁰ *Getty Images (US), Inc. v. Stability AI, Inc.*, No. 3:25-cv-06891 (N.D. Cal.) (refiled from D. Del. No. 1:23-cv-00135, voluntarily dismissed 14 Aug. 2025).

²¹ U.S. Copyright Office, *Copyright and Artificial Intelligence*, Part 3: Generative AI Training (pre-publication version, 9 May 2025).

²² Skadden, Arps, Slate, Meagher & Flom LLP, "Copyright Office Weighs In on AI Training and Fair Use" (2025); Crowell & Moring LLP, "U.S. Copyright Office Releases Third Report on AI and Copyright Addressing Training AI Models with Copyrighted Materials" (2025).

transparency-forward measure that sidesteps the harder fair use question.²³ By contrast, a broader legislative package associated with Senator Blackburn would go further and legislatively declare that unauthorised reproduction of copyrighted works for AI training is not a fair use, a position that stands in visible tension with the current administration's stated preference for a permissive training environment to preserve American AI competitiveness.²⁴ No such measure had been enacted at the time of writing, leaving the courts as the primary source of governing law.

V. The European Union

A. The Text-and-Data-Mining Exception under the DSM Directive

Unlike the United States, the European Union has a statutory exception designed with automated data analysis in mind. The Directive on Copyright in the Digital Single Market (the "DSM Directive"), adopted in 2019, created two text-and-data-mining exceptions. Article 3 permits reproductions and extractions made by research organisations and cultural heritage institutions for scientific research purposes and cannot be overridden by contract.²⁵ Article 4 is considerably broader: it permits reproductions by any person, for any purpose, including commercial purposes, subject to a critical qualification rightsholders may reserve their rights and remove their works from the exception's scope, provided the reservation is expressed in an appropriate manner, such as machine-readable means, for content made publicly available online.²⁶ Because Article 4 is not limited to non-commercial or research use, it is widely understood, including by the European Commission itself, to cover the training of commercial generative AI models by default unless a rightsholder has validly opted out. This makes the Union's baseline position considerably more permissive toward AI training than the position in the United Kingdom examined in Section VI, despite both jurisdictions sharing the Directive's text as a common origin prior to Brexit.

²³ Transparency and Responsibility for Artificial Intelligence Networks Act ("TRAIN Act"), H.R. 7209 and S. 2455, 119th Cong. (introduced 22 Jan. 2026).

²⁴ Deadline, "GOP Senator Unveils Draft Of Sweeping AI Legislation" (18 March 2026).

²⁵ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market, 2019 O.J. (L 130) 92, art. 3.

²⁶ Directive (EU) 2019/790 on Copyright and Related Rights in the Digital Single Market, art. 4.

B. Layering the AI Act on the TDM Exception

The EU AI Act does not alter the substantive TDM exception described above; instead, it imposes governance and transparency obligations on providers of general-purpose AI models placed on the EU market, regardless of where the model was developed or trained.²⁷ Under Article 53(1)(c), a provider must adopt a policy to comply with EU copyright law, in particular by identifying and complying with rights reservations made under Article 4 of the DSM Directive using state-of-the-art technology. Under Article 53(1)(d), a provider must additionally prepare and make publicly available a “sufficiently detailed summary” of the content used to train the model, in a format set by the EU AI Office, so as to assist rightsholders in enforcing their rights without requiring disclosure of trade secrets.²⁸ These obligations apply extraterritorially to any general-purpose model made available in the Union, giving the AI Act’s transparency regime a reach considerably wider than the DSM Directive’s own, more conventionally territorial, infringement provisions.

The European Commission published a voluntary General-Purpose AI Code of Practice on 10 July 2025, developed through a multi-stakeholder drafting process, which operationalises the Article 53 copyright obligations in more concrete terms: signatories commit to excluding known piracy sources from their training pipelines as a mandatory baseline measure, to maintaining a complaint-handling mechanism for rightsholders, and to sharing compliance documentation with downstream providers within a fixed period.²⁹ A separate, mandatory Training Data Summary Template was published on 24 July 2025 to standardise the disclosures required under Article 53(1)(d).³⁰ The Code’s mandatory exclusion of piracy sources mirrors, doctrinally, the piracy/lawful-acquisition distinction that emerged independently from *Bartz v. Anthropic* in the United States, a convergence discussed further in Section VIII.

C. Judicial Interpretation: *Kneschke v. LAION*

The clearest judicial test of the TDM exceptions’ application to AI training so far comes not

²⁷ 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), art. 53.

²⁸ European Commission, AI Office, “Template for the Public Summary of Training Content for General-Purpose AI Models” (24 July 2025).

²⁹ European Commission, “General-Purpose AI Code of Practice” (10 July 2025).

³⁰ European Commission, AI Office, “Template for the Public Summary of Training Content for General-Purpose AI Models” (24 July 2025).

from a claim against a commercial model developer but from an action brought by photographer Robert Kneschke against LAION e.V., a non-profit organisation that assembles and publishes large open datasets of image–text pairs used across the AI research community, including by commercial developers downstream. The Regional Court of Hamburg held in September 2024, and the Hanseatic Higher Regional Court confirmed on appeal in December 2025, that LAION’s creation of its dataset, which involved analysing Kneschke’s copyrighted photograph without permission, fell within the scientific-research TDM exception implementing Article 3, since LAION is a non-profit research organisation that made the dataset itself freely available rather than exploiting it commercially.³¹ The appellate court went further and held that the reproduction was independently justified under the general Article 4 exception as well, because the photographer’s rights reservation, expressed only in natural-language website terms rather than a machine-readable format, was not effective at the relevant time.³² The court left open, however, whether the Article 4 exception extends to every stage of a full generative-AI training pipeline, confining its holding to the comparison of images against text descriptions that precedes actual model training, and Kneschke has been granted leave to seek further review before the Federal Court of Justice.³³ No Court of Justice of the European Union ruling on the application of Articles 3 or 4 to AI training has yet been handed down, leaving Kneschke persuasive but not binding across the Union.

D. Assessment

The European Union’s approach can be summarised as a permissive substantive default commercial text and data mining, including AI training, is lawful unless a rightsholder has validly opted out paired with an increasingly demanding transparency overlay imposed through the AI Act rather than through copyright law itself. This is structurally different from the United States’ *ex post*, litigation-driven fair use model: rather than asking courts to determine after the fact whether a given training run was fair, the EU framework asks developers to comply with disclosure and rights-reservation obligations up front, leaving comparatively less room for prolonged, multi-year factual litigation of the kind seen in the American cases discussed in Section IV. Whether this yields greater legal certainty in practice remains to be tested, since

³¹ *Kneschke v LAION e.V.*, Regional Court of Hamburg, Case No. 310 O 227/23 (27 Sept. 2024).

³² *Kneschke v LAION e.V.*, Hanseatic Higher Regional Court (OLG Hamburg), Case No. 5 U 104/24 (10 Dec. 2025).

³³ *Kneschke v LAION e.V.*, Hanseatic Higher Regional Court (OLG Hamburg), Case No. 5 U 104/24 (10 Dec. 2025).

the opt-out mechanism's central term, what counts as an effective, machine-readable rights reservation, was itself genuinely disputed in Kneschke, and the interaction between Article 4's opt-out and Article 53's copyright-compliance policy has not yet been tested against a commercial foundation-model developer in litigation.

VI. The United Kingdom

A. The Existing Statutory Position

The Copyright, Designs and Patents Act 1988 ("CDPA") contains a text-and-data-mining exception at section 29A, but it is considerably narrower than either the European Union's Article 4 or the fair use approach available in the United States: it applies only where the person mining the work already has lawful access to it and does so for research for a non-commercial purpose.³⁴ No equivalent exception exists for commercial text and data mining, meaning that, on the current statutory text, training a commercial generative AI model on copyrighted works without a licence has no dedicated statutory safe harbour in the United Kingdom, a materially more restrictive default than the European Union's. The CDPA also contains an unusual provision, section 9(3), which deems the author of a computer-generated work, one generated in circumstances such that there is no human author, to be the person who made the arrangements necessary for its creation; this provision addresses the separate question of authorship in AI-assisted outputs rather than the training-input question that is this paper's focus, but it has become entangled with the broader reform debate discussed below.³⁵

B. Getty Images v. Stability AI: The First UK Judgment

On 4 November 2025, the High Court of England and Wales handed down judgment in *Getty Images (US) Inc & Ors v Stability AI Ltd*, the first UK ruling to address copyright infringement arising from the training and use of a generative AI model.³⁶ The judgment illustrates vividly how territoriality can constrain copyright enforcement against an internationally trained model. Getty's central claims, that the training and development of the Stable Diffusion model itself infringed Getty's copyright, were abandoned during trial once it became clear that Getty could

³⁴ Copyright, Designs and Patents Act 1988, c. 48, § 29A (UK).

³⁵ Copyright, Designs and Patents Act 1988, c. 48, § 9(3) (UK); Bird & Bird, "Copyright & AI in the UK: The Debate Rolls On" (2026).

³⁶ *Getty Images (US) Inc & Ors v Stability AI Ltd* [2025] EWHC 2863 (Ch) (Eng.).

not establish that any relevant training or development activity had actually taken place within the United Kingdom, since Stability's training infrastructure was located abroad.³⁷ Mrs Justice Joanna Smith accordingly never reached the question of whether UK law would treat AI training on copyrighted works as infringing, leaving the position under section 29A functionally untested even after the UK's first major AI-copyright trial. On the remaining secondary infringement claim, that Stable Diffusion itself was an "infringing copy" imported into the UK, the court held that while an intangible object can in principle be an "article" capable of infringing-copy status under the CDPA, Stability AI's model weights did not store a recognisable reproduction of Getty's photographs and therefore did not qualify.³⁸ Getty succeeded only on a narrow trademark claim, establishing that early versions of Stable Diffusion had, in a limited number of documented instances, generated outputs bearing a recognisable Getty watermark.³⁹ Getty was granted leave to appeal in December 2025, and the case's central copyright questions accordingly remain open.⁴⁰

C. A Protracted and Unresolved Policy Process

The UK Government has spent more than four years attempting, and repeatedly failing, to settle a legislative policy on AI and copyright. An initial 2022 proposal for a broad TDM exception covering commercial use was abandoned in 2023 following sustained opposition from the creative industries.⁴¹ A further consultation, launched in December 2024, proposed four options and identified as its preferred approach an Article-4-style exception permitting commercial text and data mining subject to a rightsholder opt-out; this proposal drew overwhelming opposition, with reportedly only around three per cent of more than 11,500 consultation responses supporting it.⁴² By January 2026, government ministers had publicly acknowledged that expressing a preference for the opt-out model had been a mistake.⁴³ A House of Lords Communications and Digital Committee report published on 6 March 2026 firmly rejected any commercial TDM exception, including an opt-out model, describing UK

³⁷ Paul, Weiss, "Getty Images v Stability AI: The UK Courts' First Word on Use of Copyright Works in AI Model Development" (24 Nov. 2025).

³⁸ Mayer Brown, "Getty Images v Stability AI: What the High Court's Decision Means for Rights-Holders and AI Developers" (20 Nov. 2025).

³⁹ Mason Hayes Curran, "Getty Images v Stability AI" (23 March 2026).

⁴⁰ AI Lawsuit Tracker, "Getty Images v. Stability AI: Case Status" (18 June 2026).

⁴¹ Hogan Lovells, "AI and copyright: UK outlook for 2026" (10 Feb. 2026).

⁴² Pinsent Masons, "Drop plans for AI-related copyright exception, UK ministers urged" (27 April 2026).

⁴³ Reed Smith, "UK copyright and AI report: the 'opt-out' is dead, but what comes next" (26 March 2026).

copyright law as an international “gold standard” whose real problem was not its substantive content but widespread unlicensed use and a lack of transparency by AI developers, and recommended a licensing-first approach instead.⁴⁴ On 18 March 2026, the Government published its statutory Report on *Copyright and Artificial Intelligence* and an accompanying Impact Assessment, required under sections 135 and 136 of the Data (Use and Access) Act 2025; the Report abandoned the opt-out option entirely, declined to commit to any of the four consulted options, and adopted instead a “wait and see” posture of monitoring industry-led licensing while gathering further evidence, with contemporaneous reporting indicating that substantive legislative decisions had effectively been deferred into 2027.⁴⁵ The Report also floated removing section 9(3)’s computer-generated-works protection altogether, adding a further open question to an already unsettled framework.⁴⁶

D. Assessment

The net effect of this five-year process is that the United Kingdom currently combines the most rightsholder-protective statutory default among the four jurisdictions examined, no exception at all for commercial text and data mining, with a first-instance judicial decision that could not test that default against a real training operation because of a jurisdictional and evidentiary gap, and a policy process that has, after extensive consultation, settled on maintaining the status quo indefinitely rather than resolving it. For a jurisdiction whose government has repeatedly emphasised the importance of both its creative industries and its ambitions as an AI hub, this leaves an unusual degree of legal uncertainty for both sides.

VII. India

A. The Statutory Framework: A Closed List of Fair Dealing Exceptions

India’s Copyright Act, 1957 does not adopt the open-ended, multi-factor fair use standard found in the United States.⁴⁷ Section 52 instead sets out an enumerated, closed list of acts that do not constitute infringement, among them fair dealing with a literary, dramatic, musical or

⁴⁴ Fieldfisher, “UK government maintains status quo on AI and copyright, playing the long game on potential reform” (31 March 2026).

⁴⁵ Simmons & Simmons, “A masterclass in can-kicking: UK Government views on AI and copyright” (2026).

⁴⁶ Copyright, Designs and Patents Act 1988, c. 48, § 9(3) (UK); Bird & Bird, “Copyright & AI in the UK: The Debate Rolls On” (2026).

⁴⁷ The Copyright Act, 1957, No. 14 of 1957, §§ 2(d)(vi), 51, 52 (India).

artistic work for the purposes of private or personal use, including research, and for criticism or review.⁴⁸ Unlike the position in the European Union or, prospectively, the United Kingdom, there is no exception drafted with text and data mining specifically in mind; any argument that training a generative model on copyrighted Indian works is non-infringing must therefore be built by fitting that activity within the existing “research” or “private use” language of section 52(1)(a), a considerably more improvisational exercise than applying a purpose-built exception. The Act does, however, contain a provision of independent comparative interest: section 2(d)(vi) defines the author of a computer-generated literary, dramatic, musical or artistic work as the person who causes the work to be created, a formulation that, like section 9(3) of the UK’s CDPA discussed in Section VI, contemplates authorship in works without a conventional human creator. This provision addresses who, if anyone, owns an AI-generated output rather than the training-input question that is this paper’s central concern, but it illustrates that Indian and British copyright law share a common statutory ancestry on precisely the questions AI is now testing.

B. *ANI Media v. OpenAI*: India’s First Generative AI Copyright Ruling

Until very recently, no Indian court had considered these questions in the context of generative AI. That changed with *ANI Media Pvt. Ltd. v. OpenAI Inc. & Anr.*, filed before the Delhi High Court in late 2024 by Asian News International, a major Indian news agency, which alleged that OpenAI had used its copyrighted news reports without authorisation to train ChatGPT and that ChatGPT’s outputs reproduced, and in some cases misattributed invented statements to, ANI’s reporting.⁴⁹ The case attracted significant attention both because it was India’s first such dispute and because of the range of parties it drew in: the court appointed two *amici curiae*, and the Federation of Indian Publishers, the Digital News Publishers Association and the Indian Music Industry each sought and were granted leave to intervene, reflecting the stakes the litigation held for India’s wider content industries.⁵⁰ OpenAI, for its part, argued that ChatGPT’s training took place entirely outside India, on servers it does not maintain within the

⁴⁸ IAM, “Inside India’s IP Market: a Guide 2026 The Future of Indian Copyright Legislation in the Wake of *ANI Media v OpenAI*” (25 Nov. 2025).

⁴⁹ IAM, “Inside India’s IP Market: a Guide 2026 The Future of Indian Copyright Legislation in the Wake of *ANI Media v OpenAI*” (25 Nov. 2025).

⁵⁰ Free Press Journal, “Delhi HC Refuses To Restrain OpenAI, Says ChatGPT Training On ANI Content Not Prima Facie Copyright Infringement” (24 July 2026).

country, and that the suit was accordingly not properly before an Indian court at all.⁵¹

After interim arguments concluded and the matter was reserved on 27 March 2026, Justice Amit Bansal delivered the Delhi High Court's interim order on 24 July 2026.⁵² The order is significant in two respects that bear directly on this paper's comparative analysis. First, on jurisdiction, the court ruled in ANI's favour, rejecting OpenAI's argument that the location of its training servers placed the dispute beyond the reach of Indian courts, a notably more assertive jurisdictional posture than the one taken, on materially similar facts, by the English High Court in *Getty v Stability AI*, where the claim foundered for want of evidence that training had occurred within the jurisdiction.⁵³ Second, on the merits, the court nonetheless declined to grant ANI an interim injunction, holding, on a *prima facie* basis only, that OpenAI's storage of ANI's published content for the purpose of training its models fell within the fair dealing exception in section 52(1)(a)(i) for private and personal use including research, and did not therefore amount to infringement under section 51.⁵⁴ The court further held, *prima facie*, that ChatGPT's retrieval-augmented outputs responding to queries about ANI content were not substantially similar to ANI's original reporting, and so did not independently infringe at the output stage either.⁵⁵

Two qualifications are essential to reading this ruling correctly. It is an interim order only, decided on a *prima facie* standard for the limited purpose of an injunction application, and the substantive suit continues before the roster bench to which it has now been referred; the ultimate outcome after full trial could differ.⁵⁶ It is nonetheless the first considered judicial treatment anywhere in India of whether section 52's fair dealing categories can accommodate AI training, and its willingness to read "research" in section 52(1)(a)(i) broadly enough to cover a commercial AI developer's training activity, without any legislative text-and-data-mining exception to draw on, echoes, in outcome if not in doctrinal machinery, the transformative-use reasoning that has driven the American cases discussed in Section IV, and suggests Indian

⁵¹ Free Press Journal, "Delhi HC Refuses To Restrain OpenAI, Says ChatGPT Training On ANI Content Not Prima Facie Copyright Infringement" (24 July 2026).

⁵² Adgully, "Delhi HC Denies Interim Injunction to ANI in OpenAI Copyright Suit" (24 July 2026).

⁵³ Adgully, "Delhi HC Denies Interim Injunction to ANI in OpenAI Copyright Suit" (24 July 2026).

⁵⁴ LiveLaw, "Delhi High Court Rejects ANI's Interim Plea Against ChatGPT, Says It Would Harm Public Interest" (24 July 2026).

⁵⁵ LiveLaw, "Delhi High Court Rejects ANI's Interim Plea Against ChatGPT, Says It Would Harm Public Interest" (24 July 2026).

⁵⁶ Business Standard, "OpenAI Use of ANI Data Doesn't Amount to Infringement, Says Delhi HC" (24 July 2026).

courts may be prepared to stretch existing fair dealing categories pragmatically rather than waiting for Parliament to legislate a bespoke exception.

C. The Policy and Legislative Backdrop

India's legislative posture on AI and copyright has shifted markedly even as the courts were left to address the ANI dispute without new statutory guidance. In February 2024, the ministry responsible for copyright policy took the position, in response to parliamentary questions, that the existing legal framework was adequate and that no new AI-specific category of intellectual property right was required.⁵⁷ That position had already been anticipated, and implicitly questioned, by the Parliamentary Standing Committee on Commerce's 161st Report of July 2021, which had urged the government to examine whether the Patents Act, 1970 and the Copyright Act, 1957 needed updating to address emerging technologies including AI.⁵⁸ By April 2025, the Department for Promotion of Industry and Internal Trade (DPIIT) had constituted a dedicated expert committee to examine whether the Copyright Act's existing framework was adequate to address the issues generative AI raises and to prepare a working paper for stakeholder consultation; that working paper, addressing the legal questions raised by AI training and copyright, was circulated for comment in late 2025.⁵⁹ Industry submissions to the DPIIT process, including from international software trade associations, have argued for the introduction of a clear statutory text-and-data-mining exception rather than continued reliance on courts to interpret section 52 case by case, precisely the gap the ANI interim order illustrates.⁶⁰

D. Assessment

India's position is, on paper, textually the least accommodating of the four jurisdictions toward AI training: a closed list of fair dealing categories, with no text-and-data-mining-specific exception, ordinarily leaves less interpretive room than either the United States' open-ended fair use standard or the European Union's purpose-built Article 4 exception. Yet the Delhi High

⁵⁷ Lexology, "The Indian Government on the Intersection of AI and IP Rights: No New Rules for the Moment" (1 March 2024).

⁵⁸ LexOrbis, "Recommendations of the Parliamentary Standing Committee on Commerce 'Review of the Intellectual Property Rights Regime in India'" (2021).

⁵⁹ Department for Promotion of Industry and Internal Trade, Committee on AI Systems and Copyright, Working Paper on Generative AI and Copyright (2025).

⁶⁰ Business Software Alliance, "India: BSA Submission to DPIIT Committee on Copyright and AI."

Court's interim order suggests Indian courts may be willing to achieve a functionally similar result by reading "research" and "private use" generously, at least at the interim-relief stage, while asserting jurisdiction over foreign-trained models more readily than their English counterparts have done. Whether this judicial improvisation will be confirmed at trial, and whether the DPIIT process will instead produce a dedicated statutory exception that renders the question moot, remain the two central open questions for Indian copyright law in this area.

VIII. Comparative Analysis

Table I summarises the four jurisdictions' current positions along the five axes identified in Section II

Jurisdiction	Statutory Framework	Leading Case Law (as of July 2026)	Transparency Duty	Policy Trajectory
United States	Open-ended fair use, 17 U.S.C. §107 (case-by-case)	<i>Bartz v. Anthropic</i> ; <i>Kadrey v. Meta</i> ; Ross Intelligence (on appeal); <i>NYT v. OpenAI</i> (pending)	None mandated by statute; TRAIN Act proposed	Litigation-driven; legislative outcome uncertain
European Union	Opt-out commercial TDM exception, DSM Directive art. 4	<i>Kneschke v. LAION</i> (national decision only)	Mandatory training-data summaries, AI Act art. 53(1)(d)	Substantive rule settled since 2019; transparency regime newly operative
United Kingdom	Non-commercial research TDM only, CDPA s.29A	<i>Getty v. Stability AI</i> (central claim untested on merits; on appeal)	Proposed repeatedly; not enacted	Reform paused indefinitely per March 2026 Report
India	Closed-list fair dealing, Copyright Act 1957 s.52	<i>ANI v. OpenAI</i> (interim order; trial pending)	None; DPIIT working paper in consultation	Under active DPIIT review since April 2025

Table I. Comparative summary of training-stage copyright treatment across the four jurisdictions.

Four cross-cutting observations emerge from placing these accounts side by side.

First, despite beginning from doctrinally unrelated starting points, an open-ended judicial standard in the United States, a purpose-built statutory exception in the European Union, a narrow research-only carve-out in the United Kingdom, and a closed list of fair dealing categories in India, courts in three of the four jurisdictions have converged on a broadly similar functional test: does the use serve a different purpose from the original work, and does it substitute for the original in its existing market? The American transformative-use inquiry, the Hamburg courts' willingness to treat LAION's non-commercial dataset-building as within the TDM exception's purpose, and the Delhi High Court's reading of "research" in section 52(1)(a)(i) all turn, in substance, on a similar purpose-and-market inquiry, even though they arrive at it through entirely different statutory vocabulary.⁶¹⁶²⁶³ The clearest outlier is the United Kingdom, where the absence of any commercial exception means the purpose-and-market inquiry has not yet been judicially tested at all, since Getty's claim collapsed on evidentiary grounds before the substantive question could be reached.

Second, the jurisdictions differ sharply on their default rule, which is to say, on who bears the burden of inertia while the law remains uncertain. The European Union's opt-out model places that burden on rightsholders, who must take affirmative steps to exclude their works. The United Kingdom's research-only exception places it on AI developers, who have no statutory safe harbour for commercial training absent a licence. The United States' fair use standard places it, in effect, on the courts, deciding case by case with outcomes that only become clear years after a given model was trained, as the multi-year pendency of the New York Times and Andersen litigation illustrates.⁶⁴⁶⁵ India's position is closest to the United Kingdom's on paper but, as the ANI interim order shows, may in practice be moving closer to the American model of judicial improvisation around an existing general-purpose exception.

⁶¹ *Bartz v. Anthropic PBC**, No. 3:24-cv-05417 (N.D. Cal. 23 June 2025) (order on cross-motions for summary judgment).

⁶² *Kneschke v LAION e.V.*, Hanseatic Higher Regional Court (OLG Hamburg), Case No. 5 U 104/24 (10 Dec. 2025).

⁶³ LiveLaw, "Delhi High Court Rejects ANI's Interim Plea Against ChatGPT, Says It Would Harm Public Interest" (24 July 2026).

⁶⁴ *The New York Times Co. v. Microsoft Corp. & OpenAI Inc.*, No. 1:23-cv-11195 (S.D.N.Y., filed 27 Dec. 2023); Al Jazeera, "NYT-led group asks court to sanction OpenAI in US copyright dispute" (9 July 2026).

⁶⁵ *Andersen v. Stability AI Ltd.*, No. 3:23-cv-00201 (N.D. Cal., filed 13 Jan. 2023).

Third, a genuinely convergent trend across all four jurisdictions is the emergence of the lawfulness of the source copy as an independent, increasingly decisive variable, distinct from the transformativeness of the eventual use. Judge Alsup's sharp distinction in *Bartz* between training on purchased books, which was fair use, and training on pirated books, which was not, regardless of purpose, has a direct doctrinal cousin in the European Union's GPAI Code of Practice, which makes the exclusion of known piracy sources a mandatory baseline commitment for signatories quite apart from any purpose-based analysis.⁶⁶⁶⁷ Whatever the eventual resolution of the harder transformative-use and market-dilution questions, whether the copy itself was lawfully obtained is emerging as a comparatively administrable, cross-jurisdictionally stable proxy that courts and regulators on both sides of the Atlantic are independently arriving at.

Fourth, transparency regulation is proving more tractable, politically and legally, than substantive infringement liability, and is increasingly where jurisdictions are converging even as they diverge on the underlying question of legality. The European Union's mandatory training-data summaries, the United States' pending TRAIN Act subpoena mechanism, and the House of Lords Committee's explicit preference for a transparency-and-licensing approach over any new exception in the United Kingdom all reflect a similar political logic: transparency obligations can command broader consensus than a substantive rule that necessarily favours either developers or rightsholders.⁶⁸⁶⁹⁷⁰ Whether disclosure alone can resolve the underlying distributional dispute over compensation is doubtful, however, since knowing that a work was used for training does not by itself entitle its owner to a remedy under any of the four jurisdictions' current laws.

Finally, territoriality is emerging as a significant and under-discussed strategic variable. Copyright remains, in each of the four jurisdictions, a fundamentally territorial right, yet training a generative model is an inherently transnational activity that can be, and in *Stability AI's* case was, conducted from whichever jurisdiction offers a developer the most favourable

⁶⁶ *Bartz v. Anthropic PBC**, No. 3:24-cv-05417 (N.D. Cal. 23 June 2025) (order on cross-motions for summary judgment).

⁶⁷ European Commission, "General-Purpose AI Code of Practice" (10 July 2025).

⁶⁸ European Commission, AI Office, "Template for the Public Summary of Training Content for General-Purpose AI Models" (24 July 2025).

⁶⁹ Transparency and Responsibility for Artificial Intelligence Networks Act ("TRAIN Act"), H.R. 7209 and S. 2455, 119th Cong. (introduced 22 Jan. 2026).

⁷⁰ Fieldfisher, "UK government maintains status quo on AI and copyright, playing the long game on potential reform" (31 March 2026).

legal exposure.⁷¹ The contrast between the English High Court's inability to reach Stability AI's offshore training and the Delhi High Court's willingness to assert jurisdiction over OpenAI on similar facts suggests that territoriality's practical bite will vary significantly by jurisdiction, creating both an incentive for regulatory arbitrage by developers and an argument, addressed further in Section IX, for eventual international coordination.

IX. Challenges and Policy Recommendations

The comparative picture assembled in Section VIII suggests several practical recommendations, addressed both generally and, where the analysis is most directly relevant, to India specifically.

First, jurisdictions relying on courts to resolve the underlying legality question through case-by-case litigation, principally the United States and, following the ANI interim order, potentially India, should recognise the real costs of that approach. Multi-year litigation of the kind seen in the New York Times, Andersen and Getty disputes imposes substantial costs on AI developers, who must operate for years under contingent liability, and on rightsholders, particularly smaller and individual creators who lack the resources to litigate at the scale of a wire agency or a global media company.⁷² A purpose-built statutory exception, calibrated with appropriate conditions, offers considerably more certainty than waiting for appellate courts to resolve the question one fact pattern at a time.

Second, and specifically for India, the DPIIT committee process currently underway represents an important opportunity to provide that certainty rather than leaving the question to further ad hoc judicial interpretation of section 52.⁷³ The comparative experience examined in this paper suggests two design lessons. Following the European Union's Article 4 model wholesale would resolve uncertainty quickly but would place the burden of an opt-out mechanism on Indian creators, many of whom, as the Kneschke litigation illustrates even in a sophisticated European market, may struggle with the technical requirements of an effective machine-readable rights

⁷¹ Paul, Weiss, "Getty Images v Stability AI: The UK Courts' First Word on Use of Copyright Works in AI Model Development" (24 Nov. 2025).

⁷² *The New York Times Co. v. Microsoft Corp. & OpenAI Inc.*, No. 1:23-cv-11195 (S.D.N.Y., filed 27 Dec. 2023); Al Jazeera, "NYT-led group asks court to sanction OpenAI in US copyright dispute" (9 July 2026).

⁷³ Department for Promotion of Industry and Internal Trade, Committee on AI Systems and Copyright, Working Paper on Generative AI and Copyright (2025).

reservation.⁷⁴ Following the United Kingdom's approach of prolonged non-decision, on the other hand, has demonstrably not produced legal certainty even after four years and would leave Indian creators and developers alike in the position UK stakeholders currently occupy. A calibrated middle path, a defined text-and-data-mining exception limited to training uses that do not compete with the original work's market, paired with a mandatory transparency obligation modelled on the AI Act's training-data summaries, would allow India to benefit from the legal certainty a bespoke exception provides while retaining a market-substitution safeguard for its creative and news industries, the sector whose concerns drove the ANI litigation in the first place.

Third, the emerging convergence around the lawfulness of the source copy offers a practically useful and comparatively uncontroversial starting point for any legislature drafting new rules in this area, since it does not require resolving the harder and more contested question of whether transformative training use should, in principle, be permitted at all. Excluding training data obtained through piracy or in breach of a website's terms of access is a rule most stakeholders across the AI industry and the creative industries can plausibly accept, as the EU Code of Practice's mandatory piracy exclusion and the outcome in *Bartz* both suggest.⁷⁵⁷⁶

Fourth, the territoriality problem identified in Section VIII cannot be fully solved by any one jurisdiction acting alone. A developer facing a restrictive rule in one country can, subject to practical and commercial constraints, train its models in a more permissive one and rely on the difficulty of proving where training occurred to limit its exposure elsewhere, as *Stability AI's* defence in the English proceedings illustrates.⁷⁷ This argues for eventual coordination at the level of the World Intellectual Property Organization or a similar multilateral forum, though the sharply differing policy preferences documented in this paper, the European Union's permissive opt-out default, the United Kingdom's protective non-exception, and the United States' still-unresolved litigation, make near-term multilateral consensus appear unlikely.

⁷⁴ *Kneschke v LAION e.V.*, Hanseatic Higher Regional Court (OLG Hamburg), Case No. 5 U 104/24 (10 Dec. 2025).

⁷⁵ European Commission, "General-Purpose AI Code of Practice" (10 July 2025).

⁷⁶ *Bartz v. Anthropic PBC**, No. 3:24-cv-05417 (N.D. Cal. 23 June 2025) (order on cross-motions for summary judgment).

⁷⁷ Paul, Weiss, "Getty Images v Stability AI: The UK Courts' First Word on Use of Copyright Works in AI Model Development" (24 Nov. 2025).

Finally, all four jurisdictions would benefit from developing output-stage doctrine, governing when a specific generated output substantially reproduces a specific protected work and when a pattern of stylistically similar outputs constitutes the kind of market dilution Judge Chhabria flagged in *Kadrey*, in parallel with, rather than after, the input-stage questions that have so far dominated litigation and regulatory attention.⁷⁸ As generative systems are deployed at greater scale, disputes over specific outputs, rather than the abstract legality of training, are likely to become the more frequent and practically consequential form of litigation, and none of the four jurisdictions examined in this paper has yet developed settled doctrine to address it.

X. Conclusion

No jurisdiction examined in this paper has produced a settled, stable answer to whether training a generative AI model on copyrighted works without a licence is lawful. The United States has produced significant but partial trial-court rulings, with the first appellate guidance still pending at the time of writing. The European Union has the most complete statutory architecture but has not yet seen its central provisions tested against a major commercial foundation-model developer in litigation. The United Kingdom has spent more than four years consulting on reform without legislating, while its first judicial decision addressed only a narrow secondary question. India has just produced its first substantive ruling, on an interim and provisional basis only, in a case that will not reach final judgment for some time.

Despite these unresolved postures, the comparative analysis in this paper suggests the four jurisdictions are converging, in substance if not in doctrinal form, around three regulatory ideas: a purpose-based inquiry into transformativeness that is meaningfully qualified by evidence of market substitution or dilution; growing recognition that the lawfulness of the underlying source copy is an administrable and independently significant variable; and a shift toward mandatory transparency about training data as a more politically tractable regulatory lever than resolving the underlying question of infringement liability outright. For India in particular, the DPIIT's ongoing review of the Copyright Act, 1957 offers a genuine opportunity to draw on the comparative lessons of the other three jurisdictions' more advanced, if still incomplete, experience, rather than leaving the question to further piecemeal judicial interpretation of a fair dealing framework drafted decades before anyone anticipated its

⁷⁸ *Kadrey v. Meta Platforms, Inc.**, No. 3:23-cv-03417 (N.D. Cal. 25 June 2025) (order on motion for summary judgment).

application to a machine capable of learning from the whole of the written and visual record made available to it. How that opportunity is used will shape not only the legal exposure of Indian and foreign AI developers operating in one of the world's largest AI markets, but also the practical value of copyright protection for Indian creators and news organisations in the generative AI era.