PRIVATE USE AND COMMERCIAL AI TRAINING UNDER INDIAN COPYRIGHT LAW: A CRITICAL ANALYSIS OF SECTION 52(1)(A) IN ANI MEDIA V. OPENAI'S INTERIM ORDER

Rishabh Sisodiya, National Law Institute University (NLIU), Bhopal

Samarth Udasin, National Law Institute University (NLIU), Bhopal

Priyanshu Tripathi, National Law Institute University (NLIU), Bhopal

ABSTRACT

On 24 July 2026, the Delhi High Court declined to grant Asian News International ('ANI') an interim injunction against OpenAI, holding at a prima facie stage that OpenAI's storage of ANI's news content to train the large language models underlying ChatGPT fell within the 'private or personal use, including research' limb of Section 52(1)(a)(i) of the Copyright Act 1957. This article situates that holding within three bodies of material rarely brought together in Indian scholarship: the technical mechanics of how large language models are actually trained; the comparative law of text-and-data-mining exceptions in the United States, the European Union, the United Kingdom and Japan, together with the generation of AI-training litigation those regimes have produced since 2024; and India's own, contemporaneous executive process for reforming Section 52. Read together, these materials support a structural claim: every other major regime that permits AI training on copyrighted works does so through a mechanism purpose-built to track scale: an opt-out right, a market-substitution factor, an output-side 'enjoyment' test, or a strict non-commercial limitation. Section 52(1)(a), as applied in ANI Media v OpenAI, has no independently operative version of any of these; it reached its result by routing commercial, indefinitely monetised training through a category 'private or personal use' built for an individual, capacity-bounded human researcher. The article uses the order's own technical findings and its own survey of comparative doctrine to test that routing, ahead of the pending trial and India's own live law-reform process.

Keywords: fair dealing; Section 52(1)(a); Copyright Act 1957; large language models; text and data mining; ANI Media v OpenAI; comparative copyright; AI governance.

I. Introduction

Copyright statutes are written for a narrower world than the one that eventually tests them. On 24 July 2026, that gap reached the Delhi High Court in *ANI Media Pvt Ltd v Open AI OpCo LLC*, where Justice Amit Bansal refused Asian News International an interim injunction against OpenAI, holding on a prima facie basis that OpenAI's storage of ANI's news content to train the large language models underlying ChatGPT fell within the 'private or personal use, including research' exception in Section 52(1)(a)(i) of the Copyright Act 1957.[1] It is, on the evidence available, the first substantive judicial ruling anywhere in the common-law world to apply a general fair-dealing 'private use' category as opposed to a purpose-built text-and-data-mining exception or a multi-factor fair-use test to the training of a generative large language model.

This article expands an earlier, narrower analysis of that holding into three further directions the earlier analysis did not pursue. First, it grounds the legal characterisation in the technical reality of what 'training' an LLM actually involves, tokenisation, embedding, gradient-based optimisation, and the documented phenomenon of memorisation, since the order's own findings about 'storage' and 'closed activity' are, at bottom, claims about a technical process that can be checked against the published computer-science literature and against what other courts, applying the same underlying technology to different statutory tests, have found as fact. Second, it places Section 52(1)(a) within a comparative frame: the United States' fair-use doctrine and its 2025–26 generation of AI-training decisions; the European Union's two-track text-and-data-mining exception and the German litigation testing it; the United Kingdom's narrower, non-commercial exception and its stalled reform process; and Japan's permissive Article 30-4. Third, it connects the doctrinal argument to India's own, live executive process for reforming Section 52, a process that, independently of this article, has already concluded that the provision does not comfortably reach commercial AI training.

The article's core argument, carried over and sharpened from the earlier analysis, is structural rather than result-oriented: it does not claim OpenAI must lose at trial, and it takes the order's

[1]ANI Media Pvt Ltd v Open AI OpCo LLC, IA 45300/2024 in CS(COMM) 1028/2024, IA 45301/2024 & IA 26192/2025 (Delhi High Court, 24 July 2026) (Bansal J) ('ANI Media v OpenAI'). The suit was instituted in November 2024; judgment was reserved on 27 March 2026 and pronounced on 24 July 2026. At the time of writing no neutral citation had been authoritatively confirmed; this article cites the order by case number, date and paragraph. The full text is available via the Delhi High Court's judgment portal, https://delhihighcourt.nic.in (search CS(COMM) 1028/2024), and on Indian Kanoon, https://indiankanoon.org/doc/93327052/ (both accessed August 2026).

interim, prima facie character at face value.[2] The claim is that every comparator regime examined here keeps commercial-scale AI training in check through a mechanism purpose-built to track scale: an opt-out right in the EU, a market-substitution factor in the US, an output-facing 'enjoyment' test in Japan, a blanket non-commercial limitation in the UK. Section 52(1)(a), as applied in *ANI Media v OpenAI*, has no independently operative version of any of these. It reached its result by treating 'private or personal use, including research', a category built for an individual, capacity-bounded human researcher as sufficient, on its own, to clear commercial, indefinitely monetised training. The article proceeds in nine further parts: a technical primer (Part II); the Indian framework and the order itself (Part III); comparative analysis of the United States (Part IV), the European Union (Part V), and the United Kingdom and Japan (Part VI); a synthesis situating India on this comparative spectrum, including its own pending law-reform process (Part VII); the scale-based critique of the order's own reasoning (Part VIII); an examination of the fairness test's reliance on American authority (Part IX); and a reform-oriented conclusion drawing on the comparative material (Part X).

II. The Technology of Large Language Model Training

A recurring difficulty in adjudicating AI-copyright disputes is that the legal categories, 'copy', 'storage', 'reproduction', were built around media in which the relationship between an input and a retained copy is intuitive: a photocopier makes a copy; a hard drive stores a file. The relationship between a training corpus and a trained model's parameters is not intuitive in the same way, and several of the doctrinal disputes examined in this article including the Munich Regional Court's finding in *GEMA v OpenAI* that memorisation is itself a reproduction, discussed in Part V turn on getting this relationship right. This Part sets out, at a level appropriate for legal analysis rather than for computer science, what is publicly documented about how models of this kind are built.

A. Tokenisation, Embeddings, and the Transformer Architecture

Before training begins, text is broken into 'tokens' (sub-word units) and each token is mapped to a numerical vector, an 'embedding', in a high-dimensional space designed so that tokens used in similar contexts sit closer together. The dominant architecture for large language models since 2017, the 'transformer', processes sequences of these embeddings through layers

[2]ANI Media v OpenAI (n 1) [9], setting out Issues I–IV, and the accompanying clarification that Issue III, though captioned 'fair use', is to be read as 'fair dealing', the expression used in the Copyright Act 1957, s 52(1)(a).

that repeatedly compute how much each token in a sequence should 'attend to' every other token, allowing the model to build a contextual representation of a passage rather than treating each word in isolation.[3] A trained model consists of the numerical values 'weights' or 'parameters' assigned to the connections within this architecture, typically numbering in the tens or hundreds of billions for models of ChatGPT's generation. Critically, at no point in this description does the architecture retain a copy of any training document in the form a lawyer or a librarian would recognise as a 'copy': there is no file, no searchable database entry, no direct address at which a given news article 'lives' inside the parameters.

B. Training as Gradient-Based Optimisation

Training proceeds by repeatedly asking the model, on the basis of a sequence of tokens drawn from the training corpus, to predict the next token, comparing its prediction against the actual next token in the source text, and adjusting every parameter slightly via an algorithm called backpropagation, applied through gradient descent to make that specific prediction marginally more likely next time. This step is repeated, across the entire training corpus, typically many times over. The result is a set of parameters that encode statistical regularities of language grammar, factual association, argumentative structure, stylistic register abstracted across the whole corpus, rather than a copy of any single work within it. This is the technical basis for the characterisation, accepted by courts on both sides of the Atlantic examined in this article, that training is a form of statistical abstraction rather than a form of storage in the ordinary sense: it is also the technical basis on which the Delhi High Court found that OpenAI's training data 'sits in a closed environment, inaccessible to any human user for viewing or download', a finding this article does not dispute at the level of individual documents.

C. Memorisation: When Statistical Learning Becomes Verbatim Reproduction

The abstraction described above is not, however, complete or uniform across a training corpus. A substantial body of computer-science research most influentially a 2021 study demonstrating that sequences repeated frequently, or otherwise unusual, in a training corpus can be extracted from a trained model's parameters largely verbatim through targeted prompting has

[3]A Vaswani and others, 'Attention Is All You Need' (2017) 30 Advances in Neural Information Processing Systems 5998. This paper introduced the transformer architecture underlying the publicly known LLM families, including the GPT family; OpenAI has not published full architectural or training-data specifications for the GPT-4 or GPT-4o models underlying ChatGPT, so this Part describes the transformer-based architecture publicly documented for models of this general class, not OpenAI's specific undisclosed implementation.

documented 'memorisation' as a real, empirically observed phenomenon distinct from ordinary statistical generalisation.[4] This is precisely the technical phenomenon at issue in the Munich Regional Court's November 2025 finding, discussed in Part V, that GPT-4 and GPT-4o had memorised specific German song lyrics, a finding the court reached by comparing outputs against source lyrics and ruling out coincidence given their length and complexity, and one it held constituted 'reproduction' under German and EU copyright law notwithstanding OpenAI's argument that its models 'reflect learned patterns' rather than retaining verbatim text. The same underlying technical distinction, abstraction as the general case, memorisation as a documented but comparatively rare exception, was central to the Delhi High Court's own output-claim analysis, where ANI's failure to demonstrate memorisation or regurgitation of its specific articles, rather than any finding that memorisation is technically impossible, was what defeated that limb of the claim.

D. Why These Technical Facts Matter to the Legal Characterisation

Three implications follow for the legal analysis in the remainder of this article. First, the technical distinction between abstraction and memorisation maps onto, but does not resolve, the legal distinction between the 'training claim' and the 'output claim' examined in Part III: a training process that is technically abstractive for the overwhelming majority of a corpus can still memorise particular high-frequency or distinctive passages, which is why courts examined in Parts IV and V have treated the two claims as requiring separate proof rather than assuming that a finding on one resolves the other. Second, the fact that training does not create a document-addressable 'copy' in the traditional sense is a genuine technical fact, and it is the strongest technical support for treating training as a fundamentally transformative, non-expropriative process but it is a fact about the *mechanism* of training, not about its *scale* or *commercial consequence*, and Part VIII argues that the Delhi High Court's purpose-test reasoning conflates the two. Third, and most directly relevant to the comparative analysis that follows: because the technical process is materially the same wherever in the world a model is trained, the wide divergence in legal outcomes examined in Parts IV–VI: fair use in California, infringement in Munich, an unresolved secondary-infringement question in London cannot be explained by any difference in the underlying technology. It is explained entirely by differences

[4]N Carlini and others, 'Extracting Training Data from Large Language Models' (30th USENIX Security Symposium, 2021) 2633. The paper demonstrates that sequences repeated frequently in a training corpus, or otherwise statistical outliers, can be extracted from a trained model's parameters largely verbatim through targeted prompting, a phenomenon the paper terms 'memorisation'.

in the legal mechanisms each jurisdiction uses to convert that shared technical process into a finding of lawful or unlawful use, which is exactly the comparative question this article turns to next.

III. The Indian Framework: Section 52(1)(a) and ANI Media v OpenAI

ANI sued OpenAI in November 2024, alleging copyright infringement on two counts: a 'training claim', that OpenAI copied and stored ANI's news content without authorisation to train the large language models underlying ChatGPT, and an 'output claim', that ChatGPT reproduced ANI's works, or generated fabricated statements falsely attributed to ANI, in its responses to users.[5] The Court framed four issues: (I) whether storage of ANI's data for training ChatGPT amounts to infringement; (II) whether generating user responses using that data amounts to infringement; (III) whether either use qualifies as fair dealing under Section 52; and (IV) whether Indian courts have territorial jurisdiction given that OpenAI's training occurs on servers in the United States.[6]

On jurisdiction, the Court found in ANI's favour: OpenAI's active provision of services to Indian users was sufficient to found jurisdiction under Section 62(2) of the Copyright Act and Section 20 of the Code of Civil Procedure 1908, notwithstanding that training itself occurs abroad.[7] This is a notable point of contrast with the United Kingdom's *Getty Images v Stability AI*, discussed in Part VI, where a UK court's inability to reach conduct occurring abroad forced Getty to abandon its primary claim entirely. On the output claim, the Court held that ANI had not shown memorisation or regurgitation of its literary works in ChatGPT's responses, and that the specific articles ANI relied on to demonstrate reproduction had in fact been published after the cut-off dates for the training data of the relevant model versions.[8] On the training claim, decided together with the fair dealing issue, the Court held on a prima facie basis that OpenAI's storage of ANI's literary works for training satisfied both limbs of Section 52(1)(a): the purpose test and the fairness test.[9]

[5]ANI Media v OpenAI (n 1) [6].
[6]ANI Media v OpenAI (n 1) [9].
[7]ANI Media v OpenAI (n 1) [42.1]–[42.4], [270]. Several commentators have noted that OpenAI's objection to territorial jurisdiction was in fact rejected, a point sometimes lost in headlines announcing an unqualified 'win' for OpenAI: see 'Delhi High Court Rules for OpenAI: The Third Major Fair-Use Verdict in 13 Months' The New Publishing Standard (27 July 2026).
[8]ANI Media v OpenAI (n 1) [241]–[243].
[9]ANI Media v OpenAI (n 1) [219], [256].

On balance of convenience, the Court noted that ANI had itself quantified its claim by offering OpenAI a licence for USD 7.5 million in October 2024, meaning any eventual loss was compensable in damages; and that an injunction would cause irreparable harm to millions of Indian ChatGPT users, many of them free users, as well as to the broader public interest in AI development in India.[10] The interim application was accordingly dismissed, with the Court expressly noting that its observations were confined to the interlocutory stage and would have no bearing on the final outcome of the suit.[11]

A. The Purpose Test's Three Sub-Inquiries

The Court structured its fair dealing analysis into a two-step examination: whether the storage falls within one of the purposes enumerated in Section 52(1)(a) (the purpose test), and whether that storage otherwise qualifies as fair dealing (the fairness test).[12] Within the purpose test, the Court resolved three distinct sub-questions, described briefly here as background to the comparative and critical analysis that follows.

On commerciality, the Court held that Parliament knew how to write a non-commercial limitation into Section 52(1) when it wished to and it did so expressly in clauses (ad), (k)(ii), (l), (n) and (o), and pointedly did not do so in clause (a), supporting this with the Canadian Supreme Court's holding in *CCH Canadian* that 'research' must be given a large and liberal interpretation not confined to non-commercial contexts.[13][14] On the non-infringing-copy requirement in the Explanation to Section 52(1)(a), the Court held the qualifier attaches grammatically only to computer programmes, and found support in the American decision in *Bartz v Anthropic*, discussed in Part IV, which distinguished lawfully acquired training copies from shadow-library copies; OpenAI, the Court noted, obtained ANI's content from its own publicly accessible website.[15]

On 'private or personal use, including research', the Court held that 'private', unlike 'personal', is not confined to individuals and can extend to a closed group or company, finding support in

[10]ANI Media v OpenAI (n 1) [265]–[269], recording ANI's licence offer of USD 7.5 million by communication dated 3 October 2024.
[11]ANI Media v OpenAI (n 1) [274].
[12]ANI Media v OpenAI (n 1) [168].
[13]ANI Media v OpenAI (n 1) [173]–[177], [190].
[14]CCH Canadian Ltd v Law Society of Upper Canada 2004 SCC 13, [2004] 1 SCR 339 [51], quoted in ANI Media v OpenAI (n 1) [180].
[15]ANI Media v OpenAI (n 1) [191]–[201].

the Supreme Court's decision in *B Malini Mallya*, where a dance performance by an educational institution was held capable of falling within Section 52(1)(a)(i).[16] [17] On the facts, the Court held that OpenAI's training data sits in a closed environment, inaccessible to any human user for viewing or download, rendering the use 'purely private'.[18]

On 'research', the Court adopted the dictionary meaning from *Blackwood v Parasuraman*, and made an observation central to Part VIII of this article: that research 'is normally a closed activity and is not disclosed to the general public. It is always the output of the research that is communicated to the public.'[19] Applying the 'doctrine of updating construction' from *SJ Choudhary*, the Court held that 'research' can extend to machine learning, since LLM training involves screening, organising and extracting from stored works to generate new statistical outputs.[20] It illustrated the point with an AI tutor replacing a human teacher under Section 52(1)(i), reasoning that denying the exception to an AI performing the same function would be 'regressive'.[21] On this basis, the Court concluded, prima facie, that OpenAI's training fell within 'private or personal use, including research' and satisfied the purpose test.

B. Legislative History

As originally enacted in 1957, Section 52(1)(a) protected fair dealing 'for the purposes of (i) research or private study; (ii) criticism or review'.[22] The 1994 amendment replaced 'research or private study' with 'private use, including research', explained in the Notes on Clauses as necessary because an unduly narrow reading of 'private study' had begun to cause 'harassment to public', a reference to individual researchers and students being penalised for ordinary, small-scale copying.[23] The 2012 amendment added 'personal' alongside 'private' and inserted the Explanation clarifying that electronic storage counts as protected storage on the same footing as physical copying.[24] Every amendment responds to a mischief involving an individual

[16]ANI Media v OpenAI (n 1) [202]–[219].
[17]Academy of General Education, Manipal v B Malini Mallya (2009) 4 SCC 256, AIR 2009 SC 1982 [39]–[40], quoted in ANI Media v OpenAI (n 1) [210].
[18]ANI Media v OpenAI (n 1) [211].
[19]Blackwood and Sons Ltd v AN Parasuraman AIR 1959 Mad 410, [154]–[155], quoted in ANI Media v OpenAI (n 1) [213]–[214].
[20]ANI Media v OpenAI (n 1) [216]–[217]; applying State (through CBI/New Delhi) v SJ Choudhary (1996) 2 SCC 428.
[21]ANI Media v OpenAI (n 1) [218]; Copyright Act 1957, s 52(1)(i).
[22]Copyright Act 1957, s 52(1)(a), as originally enacted; text set out in ANI Media v OpenAI (n 1) [157].
[23]ANI Media v OpenAI (n 1) [158]–[159], quoting the Notes on Clauses to the Copyright (Amendment) Bill 1992: 'an unduly narrow interpretation of the words "private study" ... may result in harassment to public'.
[24]ibid [160]–[161].

user being denied the benefit of an exception that a fair reading of legislative purpose plainly intended to cover; none contemplates an institutional actor whose 'private' processing step is the input to a persistently monetised, publicly distributed product. Part VIII returns to this history in detail; Parts IV–VI first situate it comparatively.

IV. Comparative Perspective I: The United States

The United States has no purpose-built text-and-data-mining exception; AI training there is tested, like any other use, against the open-ended four-factor fair-use standard in 17 USC §107: the purpose and character of the use (including whether it is transformative), the nature of the copyrighted work, the amount and substantiality of the portion used, and the effect on the potential market for the original. Because §107 has no purpose-based threshold comparable to Section 52(1)(a)'s enumerated categories, American courts confront the scale and commerciality of AI training directly, within a single inquiry, rather than filtering it through a prior gatekeeping category.

A. Foundational Precedent: Mass Digitisation as Transformative Use

Long before generative AI, American courts had already found mass digitisation of copyrighted works for full-text search and accessibility purposes to be fair use in *Authors Guild v HathiTrust*, which held that a searchable index built from scanned copyrighted books, which did not expose the works' expressive content to end users, was transformative notwithstanding that entire works were copied.[25] The Second Circuit reached a similar result the following year in *Authors Guild v Google*, upholding Google Books' mass scanning as fair use, and holding that a commercial motive does not, by itself, defeat a finding of transformative use.[26] These decisions supplied the doctrinal foundation: mass ingestion for a purpose distinct from consuming the works' original expressive value can be transformative even at commercial scale, on which the 2025–26 generation of AI-training cases builds.

B. The 2025–26 Generation of AI-Training Cases

The clearest illustration of §107's scale-sensitivity is that the three most significant American AI-training rulings of 2025, mechanically near-identical at the level of the training process

[25]Authors Guild v HathiTrust 755 F3d 87 (2d Cir 2014).
[26]Authors Guild v Google Inc 804 F3d 202 (2d Cir 2015), discussed in ANI Media v OpenAI (n 1) [249].

itself (Part II), reached sharply different results once the fourth factor was applied to what that process had actually produced, turning the differences into a natural experiment. In *Bartz v Anthropic*, Judge Alsup held that training Claude on lawfully purchased and scanned books was 'exceedingly transformative' and fair use.[27] But he simultaneously held that Anthropic's retention of more than seven million books downloaded from pirate libraries in a permanent digital 'central library' regardless of whether those specific copies were ever used for training was not justified by the training use and infringed separately.[28] That piracy-acquisition claim was later settled for USD 1.5 billion, the largest reported copyright recovery in history, with final court approval on 20 July 2026; the settlement releases only claims about past acquisition, not about outputs or about training on lawfully acquired material, which the court had already found lawful.[29]

In *Kadrey v Meta*, decided two days later, Judge Chhabria similarly granted partial summary judgment for Meta, finding LLM training highly transformative, but expressly warned that a stronger evidentiary record on market dilution, the risk that an LLM's outputs flood the market with material that substitutes for the works it was trained on, could change the outcome in a future case with better evidence.[30] Both courts rejected the authors' argument that a lost opportunity to license the works specifically for AI training should itself count as market harm, calling that reasoning circular since accepting it would make virtually any unlicensed transformative use infringing merely because a licence could theoretically have been charged for it.

The outlier, and the sharpest available contrast, is *Thomson Reuters v Ross Intelligence*, where Judge Bibas granted summary judgment against a legal-research start-up that had trained a competing product on Westlaw's copyrighted headnotes.[31] Unlike the training in *Bartz* and *Kadrey*, which produced general-purpose models with no direct overlap with the plaintiffs'

[27]Bartz v Anthropic PBC 787 F Supp 3d 1007 (ND Cal 2025), discussed in ANI Media v OpenAI (n 1) [199].
[28]ibid, holding that training on lawfully acquired books was 'exceedingly transformative' and fair use, but that Anthropic's retention of over 7 million books downloaded from pirate libraries in a permanent 'central library' was not justified by that training use and infringed separately.
[29]Bartz v Anthropic PBC, No 3:24-cv-05417 (ND Cal, final settlement approval 20 July 2026) (Martínez-Olguín J). Anthropic agreed to pay USD 1.5 billion, roughly USD 3,000 per work, to settle the piracy-acquisition claim for the class of works downloaded from Library Genesis and the Pirate Library Mirror; the settlement releases only claims relating to past acquisition and copying of the works through 25 August 2025, and does not resolve or release any claim concerning AI outputs.
[30]Kadrey v Meta Platforms Inc, No 23-cv-03417-VC, 2025 WL 1752484 (ND Cal 25 June 2025), quoted in ANI Media v OpenAI (n 1) [248].
[31]Thomson Reuters Enterprise Centre GmbH v Ross Intelligence Inc 765 F Supp 3d 382 (D Del 2025) (Bibas J, sitting by designation).

own markets, Ross's training was aimed squarely at building a substitute for Westlaw itself; the court held the use non-transformative and found the fourth factor, market effect, which it called 'the single most important element of fair use', decisive against Ross.[32] The case is currently on appeal to the Third Circuit, argued in mid-2026, with the panel pressing both sides on how far a market-substitution finding in this context should extend.

The case most directly analogous to *ANI Media v OpenAI* on its facts is a news organisation suing OpenAI over training on its journalism remains unresolved on the merits. In *The New York Times v OpenAI*, filed in December 2023, Judge Stein's April 2025 opinion allowed the core infringement claims to proceed past the pleading stage, holding that fair use could not be decided without a developed factual record; the case reached cross-motions for summary judgment in 2026 without a ruling on the fair-use defence itself.[33] No American court has yet ruled, on a developed record, on whether training a general-purpose LLM on a news publisher's copyrighted journalism, the precise question *ANI Media v OpenAI* answered on an interim record, is fair use.

C. What the American Trilogy Reveals

Read together, *Bartz*, *Kadrey* and *Ross* show that American law's answer to the scale problem is not a threshold category at all, but a single, always-applicable factor: does the trained model, or the product built on it, substitute for the market the plaintiff's work competes in? Training that produces a general-purpose model with no direct market overlap with the source material has, on this record, consistently been found transformative regardless of its commercial scale; training that produces a direct competitor to the source material's own market has, with equal consistency, not been. Every American case discussed here reaches the market-substitution question on the merits, in every case, because §107 has no analogue to Section 52(1)(a)'s 'private or personal use' gateway that could dispose of the inquiry before that question is ever asked. This is the first of the four comparative mechanisms referred to in Part I: scale is tracked

[32]ibid, holding that Ross's use was not transformative because it 'took the headnotes to make it easier to develop a competing legal research tool', and that the fourth factor -- effect on the market, including the potential market for licensing content as AI training data -- weighed decisively against Ross because its product was intended as a direct market substitute for Westlaw.

[33]The New York Times Co v Microsoft Corp and OpenAI Inc, No 1:23-cv-11195 (SDNY, filed 27 December 2023) (Stein J). The court's 4 April 2025 opinion largely denied the defendants' motion to dismiss, holding that direct and contributory copyright infringement claims were adequately pleaded and that the fair-use defence could not be resolved on the pleadings; as of mid-2026 the case was in cross-motions for summary judgment, with no ruling yet on the merits of the fair-use defence.

through an output-facing market test applied without exception, not through a purpose-based category that can be satisfied as *ANI Media v OpenAI* found it was before market effects are considered at all.

V. Comparative Perspective II: The European Union

A. The DSM Directive's Two-Track Text-and-Data-Mining Exceptions

Unlike India and the United States, the European Union has, since 2019, addressed AI-relevant text-and-data-mining directly, through two distinct exceptions in the Directive on Copyright in the Digital Single Market.[34] Article 3 provides a narrow, mandatory exception permitting reproduction for text-and-data-mining carried out by research organisations and cultural heritage institutions for scientific research, which cannot be contracted around. Article 4 provides a broader exception, available to any person or entity for any purpose expressly including commercial AI training over lawfully accessible works, but subject to a critical qualification: rightholders may reserve their rights, opting their content out of the exception, through machine-readable means such as metadata or a website's technical protocols.[35] Article 4 is, in other words, the closest thing in any jurisdiction examined in this article to a direct legislative answer to the scale problem this article identifies in Section 52(1)(a): the exception applies by default at any scale, but the scale and commercial character of a given developer's use is precisely what determines whether a rightholder is likely to have bothered to opt out, and the opt-out right gives rightholders, not courts applying an inherited nineteenth-century category, the power to reserve their rights prospectively and technically, in advance of any specific training run, rather than a power exercised retrospectively once a use's commercial consequences have already become apparent.

B. The AI Act's Overlay: Article 53

Since August 2025, the EU AI Act has layered a transparency and compliance overlay onto the DSM Directive's substantive exceptions. Article 53 requires providers of general-purpose AI models to put in place a policy to comply with EU copyright law in particular, to identify and honour Article 4(3) rights reservations, including through automated, state-of-the-art detection

[34]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] OJ L130/92, arts 3-4 ('DSM Directive').
[35]ibid art 4(1), (3). Article 4(3) permits rightholders to reserve their rights 'in an appropriate manner, such as machine-readable means in the case of content made publicly available online' (recital 18).

of machine-readable opt-outs and to publish a sufficiently detailed public summary of the content used to train the model, using a template issued by the EU's AI Office.[36] No comparator jurisdiction examined in this article, including India, imposes an equivalent ex ante transparency obligation tied directly to copyright compliance; the closest Indian analogue, discussed in Part VII, exists only as a proposal within the DPIIT's pending consultation.

C. Germany Tests the Directive: Two Contrasting Rulings

The DSM Directive's two exceptions have now been tested in litigation, with contrasting results that illustrate how much interpretive work remains even within an express statutory TDM regime. In *LAION v Kneschke*, a photographer sued the non-profit AI-research organisation LAION over the inclusion of one of his images, scraped from a stock-photo website, in LAION's publicly released training dataset. The Hamburg Regional Court held LAION's use fell within Article 3's research exception, rejecting the argument that LAION's dataset being used downstream by commercial AI developers disqualified LAION itself, a non-profit making the dataset freely available, from the exception.[37] On appeal, the Hanseatisches Oberlandesgericht confirmed the result in December 2025, additionally holding that a rights reservation expressed only in a website's ordinary-language terms of use as opposed to a machine-readable technical protocol did not meet Article 4(3)'s standard during the period in question, a ruling of considerable practical significance for rightholders seeking to rely on the opt-out.[38]

The more consequential ruling for the argument of this article is *GEMA v OpenAI*, in which Germany's music-rights collecting society sued OpenAI over the memorisation and near-verbatim reproduction of nine well-known German songs' lyrics by ChatGPT.[39] The Munich Regional Court found, as a technical matter consistent with the memorisation literature discussed in Part II, that the lyrics were reproducibly contained within the model's parameters

[36]Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [2024] OJ L1689, art 53(1)(c)-(d) ('EU AI Act'). The copyright-related obligations in art 53 entered into force on 2 August 2025.

[37]LAION eV v Kneschke, Landgericht Hamburg, 27 September 2024, Az 310 O 227/23, holding that LAION's creation of the LAION-5B training dataset by web scraping fell within the DSM Directive art 3 / Sec 60d UrhG exception for text-and-data-mining by a research organisation for scientific research.

[38]Kneschke v LAION eV, Hanseatisches Oberlandesgericht Hamburg, appeal judgment, December 2025, Az 5 U 104/24, affirming that the reservation of rights relied on by the claimant, expressed only in the natural-language terms of a stock-photo website, was not 'machine-readable' within DSM Directive art 4(3) during the relevant period of use in 2021, and further confirming that the reproduction was justified under the art 3 research exception.

[39]GEMA v OpenAI Ireland Ltd and OpenAI OpCo LLC, Landgericht München I, 11 November 2025, Az 42 O 14139/24.

inferred by comparing training-data lyrics against model outputs and ruling out coincidence given their length and complexity and held that this memorisation itself constituted an act of reproduction under German and EU law, regardless of OpenAI's argument that the content existed only as statistical probability values rather than as a stored file. Critically for this article's purposes, the court held that the DSM Directive's text-and-data-mining exceptions cover only the *initial analytical phase* of training, the process of screening and extracting statistical patterns, and do not extend to protect the resulting memorisation once it crystallises within the model, still less its subsequent public reproduction on demand.[40] This is a materially narrower reading of what a TDM-type exception can absorb than the Delhi High Court's treatment of 'private or personal use, including research' as covering the training process in its entirety: Munich drew a line between the analytical act and its memorised residue that Delhi's single, undifferentiated finding of 'purely private' use for the whole training process did not draw. OpenAI is expected to appeal the Munich ruling; no German appellate decision had been reported as of the date of this article.

D. The EU Model's Built-In Scale Mechanism

The German litigation shows that Article 4's opt-out is not a panacea, its machine-readability requirement has already generated a full round of first-instance and appellate litigation over what counts as an adequate reservation but it demonstrates a structural feature absent from Section 52(1)(a) altogether: the EU regime places the power to respond to a use's growing scale and commercial consequence in the hands of the rightholder, exercised prospectively and technically, rather than leaving a court to infer, after the fact and on a single application for interim relief, whether an already-deployed, revenue-generating product should be redescribed as something other than what its own record shows it to be. Article 53 AI Act reinforces this by requiring AI developers themselves to build the technical capacity to detect and honour such reservations. Nothing resembling this mechanism exists in the Indian framework as applied in *ANI Media v OpenAI*, a point developed further in Part VII.

[40]ibid, holding that the memorisation of the lyrics of nine German songs within the parameters of OpenAI's GPT-4 and GPT-4o models constituted 'reproduction' under s 16 UrhG and art 2 of the InfoSoc Directive 2001/29/EC, irrespective of the fact that the content exists in the model only as probability values distributed across many parameters rather than as a discrete file; and that the text-and-data-mining exceptions in ss 44b, 60d UrhG (implementing DSM Directive arts 3-4) apply only to the initial analytical phase of training and do not extend to the resulting memorisation, still less to ChatGPT's subsequent reproduction of the memorised lyrics in response to user prompts.

VI. Comparative Perspective III: The United Kingdom and Japan

A. The United Kingdom: A Narrow Exception and a Stalled Reform

The United Kingdom's text-and-data-mining exception, Section 29A of the Copyright, Designs and Patents Act 1988, is materially narrower than either the EU's Article 4 or India's judicially extended Section 52(1)(a): it permits copying for computational analysis only where the user has lawful access to the work and the analysis is for the sole purpose of non-commercial research, a restriction that excludes essentially all commercial AI training on its face.[41] The UK government has, since 2022, run successive consultations on whether to broaden this exception including a 2024–2025 consultation, drawing over 11,500 responses, that initially favoured an EU-style opt-out model but its March 2026 Report on Copyright and Artificial Intelligence abandoned that preferred option in the face of sustained opposition from the creative industries, leaving no exception broader than Section 29A on the table and no alternative proposal endorsed in its place.[42] The result, as of the date of this article, is that commercial AI training in the UK sits in a licensing-only default position considerably more restrictive than India's post-*ANI Media* position, the opposite of what a naive assumption that common-law systems converge might predict.

The UK's only substantive judicial ruling on AI training and copyright to date, *Getty Images v Stability AI*, illustrates how much a jurisdictional gap can leave unresolved.[43] Getty was forced to abandon its primary claim that training Stable Diffusion on its images was itself infringing because it could not show that the training occurred within the United Kingdom, leaving the substantive question of training-stage infringement wholly undecided by any UK court. On the surviving secondary-infringement claim, the High Court held that Stable Diffusion's model weights contain no retrievable reproductions of Getty's images and are therefore not an 'infringing copy' imported into the UK, a finding broadly consistent with this article's Part II description of training as statistical abstraction rather than storage, but one reached without ever engaging the market-substitution or purpose-based analysis that did the decisive work in

[41]Copyright, Designs and Patents Act 1988, s 29A (inserted by the Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014, SI 2014/1372, in force 1 June 2014).
[42]UK Department for Science, Innovation and Technology, Intellectual Property Office and Department for Culture, Media and Sport, Report on Copyright and Artificial Intelligence (18 March 2026), published under the Data (Use and Access) Act 2025, ss 135-136.
[43]Getty Images (US) Inc and others v Stability AI Ltd [2025] EWHC 2863 (Ch) (Joanna Smith DBE).

the American and Indian cases.[44] Getty has indicated it will rely on the UK findings of fact in its ongoing, parallel US litigation against Stability AI, underscoring how thoroughly these disputes now cross jurisdictional lines.

B. Japan: The Permissive Outlier and Its 'Non-Enjoyment' Limit

Japan sits at the opposite end of the comparative spectrum from the UK. Article 30-4 of the Japanese Copyright Act, broadened in 2018, permits the use of copyrighted works for commercial or non-commercial purposes, with no opt-out mechanism wherever the purpose is not to 'enjoy' the thoughts or sentiments the work expresses, expressly including data analysis and machine learning, which has led commentators to describe Japan as a 'machine learning paradise'.[45] The Japanese Copyright Office's subsequent guidance has narrowed this considerably in practice: the exception does not apply where an 'enjoyment purpose' coexists with the analytical purpose for instance, fine-tuning techniques deliberately designed to let a model reproduce a specific artist's recognisable style and the proviso against 'unreasonably prejudicing' a rightholder's interests independently excludes training on content protected by technical access-control measures.[46] Japan's mechanism for tracking scale and consequence is accordingly output-facing rather than input-facing: the exception is unconditional at the point of training, but withdraws the moment a model's *output* behaviour not the training process itself shows that copyrighted expression, rather than mere statistical pattern, is what the model has been built to reproduce. This is structurally close to the distinction Munich drew in *GEMA v OpenAI* between the analytical phase and memorised residue, arrived at through a very different statutory route.

VII. Situating India on the Comparative Spectrum

Four distinct mechanisms recur across the jurisdictions examined in Parts IV–VI, each

[44]ibid, holding that Stable Diffusion's model weights do not contain or store reproductions of the training images and are accordingly not an 'infringing copy' for secondary infringement purposes, while leaving open, for want of evidence that training occurred in the UK, whether training itself would constitute primary infringement.

[45]Copyright Act (Japan, Act No 48 of 1970) art 30-4, as amended in 2018, providing that it is permissible to exploit a work to the extent necessary where it is not a person's purpose to cause a person to 'enjoy the thoughts or sentiments expressed in that work', including for the purpose of data analysis, 'provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner'.

[46]Agency for Cultural Affairs (Japan), 'General Understanding on AI and Copyright in Japan' (2024-25), clarifying that art 30-4 does not apply where an 'enjoyment purpose' coexists with the analytical purpose, in particular where training intentionally targets the reproduction of a specific work's original expression -- for example certain fine-tuning techniques designed to reproduce a particular artist's style -- and that the proviso against 'unreasonably prejudicing' rightholders' interests independently excludes training on works where the rightholder has deployed technical measures, such as a robots.txt instruction, to prevent automated collection.

performing the same underlying function: keeping a legal permission to train on copyrighted works answerable to the scale and commercial consequence of that training, rather than letting permission attach once and for all at the moment training begins. The United States uses an ever-present, output-facing market-substitution factor within §107, applied without exception in every case regardless of how the threshold 'purpose and character' factor is resolved. The European Union uses a rightholder-controlled opt-out, exercised prospectively and technically, reinforced by the AI Act's transparency obligations. The United Kingdom uses a blanket non-commercial limitation that excludes large-scale commercial training from the exception's scope entirely, leaving it to licensing markets or future legislation. Japan uses an output-facing 'non-enjoyment' test that permits training unconditionally but withdraws permission the moment a model's behaviour, not its training process, shows it has been built to reproduce rather than merely to learn from protected expression.

Section 52(1)(a), as applied in *ANI Media v OpenAI*, has no independently operative version of any of these four mechanisms. It reached its result through a fifth, structurally different route: a purpose-based threshold category, 'private or personal use, including research', that is satisfied, however elastically, through the interpretive extension traced in Part III-A, without any scale-sensitive standard built into the category itself: no opt-out, no non-commercial limitation, no output-facing withdrawal condition, and, as Part IX shows, no market-substitution check that is not itself borrowed, after the category is satisfied, from foreign case law the order's own comparative survey elsewhere treats as lacking persuasive value on the applicable domestic test. This is not merely this article's own assessment. India's Department for Promotion of Industry and Internal Trade published a Working Paper on Generative AI and Copyright in December 2025 while *ANI Media v OpenAI* was under reserved judgment whose own analysis of Section 52 concludes that 'commercial AI training is unlikely to fall within those enumerated exceptions', and records that India, unlike the United Kingdom, Japan or the European Union, has 'no general text-and-data-mining exception' at all.[47] The Working Paper's own preferred solution was a mandatory blanket licence for AI training administered through a proposed collecting society; that it also records a dissenting industry recommendation favouring an EU-style opt-out instead is itself an implicit acknowledgment that Section 52(1)(a) was never designed to do the gatekeeping work *ANI Media v OpenAI* has

[47]DPIIT, Ministry of Commerce and Industry (Government of India), Working Paper on Generative AI and Copyright (December 2025), s 2.3, available at https://www.dpiit.gov.in (consultation closed 6 February 2026).

now asked it to do.

The Ministry of Electronics and Information Technology's India AI Governance Guidelines, published the same month, reach a materially identical conclusion from the executive rather than the industry side, noting that existing copyright law 'may need to be amended' to enable large-scale AI training while protecting rightholders.[48] Taken together, the DPIIT Working Paper and the MeitY Guidelines show two arms of the Indian executive concluding, independently of this article and substantially contemporaneously with *ANI Media v OpenAI*, that the very gap the judgment filled through interpretation is one Parliament itself may need to address through legislation precisely the path the EU took in 2019, Japan in 2018, and the UK first in 2014 and then, more ambivalently, through its unresolved 2024–2026 consultation. Part VIII now returns to the order's own reasoning to show, independently of this comparative and institutional context, why its chosen route through 'private or personal use' sits uneasily with the provision's own history and the order's own later findings.

VIII. The Scale Problem: Testing the Order's Reasoning Against Itself

A. Proportionality as an Unstated Premise

The Court's own description of why research counts as 'private' is worth revisiting in light of Part II's technical primer: research, the Court held, 'is generally an intermediary process in all cases. It is undertaken before an output is generated ... It is normally a closed activity and is not disclosed to the general public. It is always the output of the research that is communicated to the public.'[49] This formula is accurate as a description of an individual researcher: a single human being's finite capacity mediates between the private input and the public output, so the scope of the 'private' phase and the scope of the eventual public output are bound together by the same bottleneck. As Part II showed, LLM training removes exactly that bottleneck: the trained model is a general-purpose engine capable of generating an effectively unlimited stream of outputs to an effectively unlimited number of users, indefinitely, at a marginal cost that falls with scale. The proportionality that made the 'closed input, public output' structure a workable

[48]Ministry of Electronics and Information Technology (Government of India), India AI Governance Guidelines (November 2025), noting that 'existing laws on copyright may need to be amended, for example, to enable large-scale training of AI models while ensuring adequate protections for copyright holders and data principals'.
[49]ANI Media v OpenAI (n 1) [214].

definition of privacy for a human researcher does not survive the substitution of a machine whose whole commercial point is to decouple the volume of output from any human bottleneck.

This is, it should be stressed, an inference this article draws from the order's own descriptive language, not a premise the order itself purports to state or defend; the argument does not depend on the Court having consciously assumed a single-researcher scale, only on the observation that its formula, read literally, supplies no textual basis for distinguishing a human researcher from an AI training pipeline once both satisfy the same 'undisclosed to the general public' description.

B. The Updating-Construction Doctrine's Limits

The Court's reliance on the doctrine of updating construction, via *SJ Choudhary*, is doctrinally sound as a general matter, but the analogy is instructive for what it shares with, and lacks in common with, the present case. Typewriting analysis was held to fall within 'science' in Section 45 of the Evidence Act 1872 because it performs precisely the same evidentiary function as handwriting analysis, at the same forensic scale: one expert, one document, one court proceeding. Extending 'private or personal use, including research' to LLM training does something different in kind: it multiplies the practical reach of the exception by orders of magnitude, because the product of the 'research' is not a single downstream work but a persistent commercial service consumed by millions of users on a continuing, revenue-generating basis. The Court's own example of an AI tutor replacing a human teacher under Section 52(1)(i) illustrates the same slippage: the teacher exception is bounded by its own text to reproduction 'in the course of instruction' within a defined class, of a defined size so an AI tutor operating within an equivalent bounded classroom would raise no scale objection at all, but ChatGPT is not that AI tutor. It is offered to the public generally, without enrolment and without the pedagogical containment that gives the human-teacher exception its textual limit.

C. The Balance-of-Convenience Findings Confirm the Scale Objection

The strongest evidence that the order's own record already registers the scale problem comes from its balance-of-convenience findings. There, the Court declined to grant an injunction precisely because ChatGPT is used by 'millions of users in India', many of them free users; because AI 'has completely revolutionized the manner in which people seek and obtain information'; and because requiring an LLM developer to obtain licences from every data

source before training 'would be economically unviable'.[50] Every one of these findings is a finding about scale. They are, if anything, findings that the end product of the training process the Court called 'private' a few dozen paragraphs earlier is about as public, and about as economically consequential, as a digital product can be in contemporary India a tension the American market-substitution factor, the EU's opt-out, and Japan's output-facing test are all, in their different ways, designed to register at the point of decision rather than only at the point of remedy.

IX. The Fairness Test's Borrowed Engine

One might respond that even if the purpose test is stretched, the fairness test operates as an independent safeguard and the Court treats the two as cumulative requirements.[51] On inspection, the fairness test as applied does not supply independent, India-specific ballast; it substantially reproduces reasoning drawn from precisely the American apparatus the order's own survey of Indian precedent treats with real ambivalence.

Surveying Indian case law, the Court recorded that the Division Bench in *Rameshwari Photocopy Services* had specifically held that decisions of the US courts applying the four-factor test 'would have no persuasive value in the Indian context', since Indian fair dealing has no equivalent codified framework.[52] The Court went further, recording 'broad consensus amongst the counsel in the present case also' that the US four-factor test is not applicable in India.[53] On that basis, the Court formulated three fairness factors of its own: whether OpenAI's use was limited to training; whether it caused economic competition or damage to ANI; and whether it served the public interest.[54]

Yet on the second factor, the Court's reasoning draws directly on, and quotes at length, the market-substitution and transformative-purpose analysis of the very American decisions discussed in Part IV, decided squarely under the four-factor framework it had just described as carrying no persuasive value in India: *Bartz v Anthropic*, which found training 'quintessentially

[50]ANI Media v OpenAI (n 1) [266]–[269].
[51]ANI Media v OpenAI (n 1) [168].
[52]The Chancellor, Masters and Scholars of the University of Oxford v Rameshwari Photocopy Services (2016) SCC OnLine Del (DB), discussed in ANI Media v OpenAI (n 1) [228]–[229].
[53]ANI Media v OpenAI (n 1) [234].
[54]ANI Media v OpenAI (n 1) [235]–[238].

transformative',[55] a parallel finding in *Kadrey v Meta*,[56] and *Authors Guild v Google*'s holding that commercial motive does not defeat transformative use.[57] The Court's own finding that OpenAI's use 'would not result in market substitution of ANI's works' follows immediately from this American authority.[58]

The tension this creates is real, and Part IV sharpens it further: the American market-substitution factor that the Delhi High Court borrows for its second fairness factor is, in the jurisdiction of origin, an *unconditional, ever-present* check applied in every case including *Ross Intelligence*, where it defeated fair use entirely. Imported into the Indian framework, the same reasoning is applied only *after* the purpose test has already, and independently, cleared the use as 'private'. Either the transformative-use reasoning of *Bartz*, *Kadrey* and *Authors Guild* has genuine interpretive purchase in the Indian statutory scheme in which case the position recorded in *Rameshwari Photocopy Services* was overstated or it does not, in which case the fairness test's market-substitution finding rests on authority the order's own survey treats as without persuasive value. Either way, the fairness test cannot be assumed to supply a genuinely independent, India-specific check on the scale-blindness identified in Part VIII; substantially the same American reasoning does significant work in both places, without the American factor's own unconditional, gateway-free application.

A defender of the order might resist the dichotomy just drawn: a court can disclaim a foreign multi-factor test as the governing legal standard while still finding specific foreign reasoning useful on a narrow, largely factual sub-question nested inside a differently structured domestic test, without thereby conceding that *Rameshwari Photocopy Services* overstated anything. That middle position is a familiar and generally legitimate feature of comparative reasoning, and it is available to the order in principle. But the order does not, at the point it relies on *Bartz*, *Kadrey* and *Authors Guild*, stake out that middle position or explain why its reliance there is compatible with the more categorical language of its own earlier survey. Absent that explanation, the tension identified above stands: not because the middle position is unavailable in the abstract, but because the order itself does not use it.

[55]Bartz v Anthropic PBC (n 27), quoted in ANI Media v OpenAI (n 1) [247].
[56]Kadrey v Meta Platforms Inc (n 30), quoted in ANI Media v OpenAI (n 1) [248].
[57]Authors Guild v Google Inc (n 26), discussed in ANI Media v OpenAI (n 1) [249].
[58]ANI Media v OpenAI (n 1) [250]–[251].

X. Toward a Reform Path: Lessons from Comparative Practice

None of the foregoing requires abandoning the order's overall framework, and the trial court, or a Division Bench should ANI appeal within the limitation period, need not discard the purpose/fairness structure, the rejection of a blanket non-commercial limitation, or the finding that electronic storage is protected on the same footing as physical copying.[59] A more conservative judicial correction is available within the order's own architecture and its own act-based structure: courts could preserve 'private or personal use, including research' for the internal, non-publicly-disclosed storage step under section 14(a)(i), while requiring the sustained commercial deployment of the resulting model, already pleadable, as this case shows, as a distinct claim for communication to the public under section 14(a)(iii), the order's own 'output claim', to carry a properly independent, scale-sensitive fairness inquiry that does not turn solely on proof of memorisation. This mirrors, in substance, the line Munich drew in *GEMA v OpenAI* between the analytical training phase and its memorised, publicly reproducible residue, and the line Japan draws between training and 'enjoyment'-oriented output, save that neither of those regimes needed to improvise the distinction judicially, because their statutes already draw it.

A legislative correction, of the kind India's own DPIIT Working Paper and MeitY Guidelines already contemplate, has the comparative record of Parts IV–VI to draw on. The DPIIT's majority-preferred 'Hybrid Model', a mandatory blanket licence administered through a proposed collecting society, with royalties triggered only once an AI product generates revenue, avoids the scale problem by making payment, rather than a court's characterisation of the input as 'private', track commercial consequence directly; it resembles no single comparator examined here as closely as it resembles a compulsory-licensing solution to the EU's opt-out problem, guaranteeing rightholders compensation without requiring the technical machine-readability apparatus that generated a full round of litigation in *Kneschke v LAION*. The dissenting industry proposal within the same Working Paper, favouring a two-tier opt-out, would import the EU's Article 4 mechanism more directly, with the accompanying risk visible in the German litigation that the adequacy of any given opt-out becomes its own extended source of dispute. A third option, not currently on the table in India but visible in the American

[59]Under the Commercial Courts Act 2015, s 13, read with the Code of Civil Procedure 1908, Order XLIII, an order refusing an interim injunction under Order XXXIX is appealable to a Division Bench; the sixty-day limitation period had not expired as of the date of this article. See also ANI Media v OpenAI (n 1) [274].

record, would be to make an explicit market-substitution inquiry of the kind that separated *Bartz* and *Kadrey* from *Ross Intelligence*, a mandatory, unconditional component of the Section 52(1)(a) fairness test, closing the gap this article has identified without waiting for Parliament to legislate a dedicated TDM exception at all.

Whichever path India takes, the comparative record assembled in this article converges on a single point: no jurisdiction that has thought carefully about AI training and copyright has left the scale question to be answered, once and for all, at the threshold, by a purpose-based category built for an individual human researcher. The United States checks it on the output side, in every case, without exception. The European Union checks it through a right the rightholder controls prospectively. Japan checks it through the character of the output rather than the character of the input. The United Kingdom, for now, declines to check it at all by simply excluding commercial training from its exception's scope. India's own executive branch has already concluded, independently of the courts, that Section 52(1)(a) needs help doing this work. *ANI Media v OpenAI* is the first judicial attempt to make the existing provision do it unaided; Part VIII of this article has tried to show, from the order's own reasoning, why that attempt strains against the very history and structure of the provision it applies.

XI. Conclusion

ANI Media v OpenAI is, on its own express terms, an interim and prima facie word, not a final one. This article has tried to test one analytical step inside a much longer order against three bodies of material the order itself does not engage in detail: the technical mechanics of what LLM training actually does to a text corpus; the comparative record of how the United States, the European Union, the United Kingdom and Japan have each built scale-sensitive mechanisms into their own AI-training doctrine; and India's own, contemporaneous executive recognition that Section 52 was not built for this task. None of this compels the conclusion that OpenAI must lose at trial, or that Section 52(1)(a) is incapable of accommodating AI training in some form: American, European and Japanese law all accommodate it, in structurally different ways. It does suggest that the specific route the Delhi High Court chose, folding an unbounded commercial pipeline into a purpose-based category built for a bounded human researcher, without an independently operative version of the opt-out, market-substitution, non-commercial, or output-facing safeguards every comparator jurisdiction employs, asks three familiar words to do work that neither their drafting history, nor the order's own later

findings, nor the comparative and technological record assembled here, comfortably supports. Whether the trial court, a Division Bench, or Parliament, already at work on the DPIIT's Working Paper, closes that gap first remains to be seen; what this article has tried to show is that the material for closing it is already on the record, at home and abroad.