# SAME DATA, DIFFERENT RULES: COPYRIGHT, DATA PROTECTION, AND THE TRAINING OF GENERATIVE AI MODELS IN THE UNITED STATES, THE EUROPEAN UNION, AND INDIA

Akshat Rawal, Manav Rachna University

**ABSTRACT**

Generative AI systems are commonly trained on very large collections of online material. Those collections may contain both copyrighted works and personal data, bringing the same training activity under two areas of law. Copyright law governs the use of protected works; data-protection law governs the processing of information about people. Each has its own rules on permission, exceptions, and remedies.

This article compares the United States, the European Union, and India. It argues that copyright and data-protection requirements operate cumulatively: meeting one does not establish compliance with the other. The two fields intersect most clearly when developers assess where training data came from and whether a model can retain or reveal information from that data. The article also examines licensing proposals, particularly India's proposed mandatory blanket licence. It argues that licensing may address disputes over copyrighted works, but cannot by itself resolve privacy obligations owed to individuals. The analysis distinguishes established law from interpretation and flags authorities that require further verification.

**Keywords:** generative AI; text and data mining; fair use; GDPR; Digital Personal Data Protection Act; training data; memorization; licensing; machine-readable opt-out.

## I. Introduction, Research Question, and Thesis

In June 2025, two judges in the Northern District of California ruled, based on the records before them, that using copyrighted books to train large language models amounted to fair use.[1] In September 2026, the Third Circuit appears to have reached the first federal appellate ruling on a related issue, finding that copying Westlaw headnotes to train a competing legal-search product was not fair use.[2] Between those decisions, a Hamburg appellate court found that creating an image-text dataset fell within the EU's text-and-data-mining exception;[3] a Munich court found that a model had reproduced memorized song lyrics;[4] and the European Data Protection Board (EDPB) cautioned against assuming that a model trained on personal data is anonymous.[5]

These decisions address copyright or privacy separately. Few consider both, even though the underlying activity—feeding large collections of text and images into a model—is the same.

Research question. When generative AI training material is both copyrighted and personal, how do copyright and data-protection laws interact in the United States, the European Union, and India? Does complying with one regime make compliance with the other easier or harder?

Thesis. The two regimes apply cumulatively; neither replaces the other. First, a copyright exception, fair-use finding, or licence does not establish a lawful basis for processing personal data, just as privacy compliance does not resolve copyright liability. Second, both regimes focus on the provenance of training data and on whether a model memorizes it, but they draw different legal consequences from those facts. Third, licensing approaches increasingly used to resolve copyright disputes cannot discharge duties owed to individuals. Proposals that treat licensing as a complete solution risk leaving data subjects without protection.

### A. Why jurisdiction matters

Copyright is territorial. Whether copying during training is lawful depends in part on where it occurs and which exceptions apply there. Data-protection rules can reach beyond national borders. The GDPR, for example, applies to controllers outside the EU that target or monitor people in the EU,[7] while the EU AI Act's recitals say that providers of general-purpose AI models must respect EU copyright law even when training takes place elsewhere.[6] India's DPDP Act can also apply to processing outside the country when it relates to offering goods

or services to people in India.[8]

The three jurisdictions illustrate different approaches. U.S. copyright law has an open-ended fair-use defence. EU law provides defined text-and-data-mining (TDM) exceptions, including an opt-out for some uses. India has a limited list of fair-dealing purposes and no TDM exception. Their privacy frameworks also differ: the United States has no comprehensive federal privacy statute, the EU has an established rights-based system, and India has a newer consent-centred law whose main duties are not yet fully in force.

Assumptions. This article focuses on training generative models using scraped or otherwise acquired text and images. It does not examine output liability, AI authorship, or biometric systems in detail, except where relevant. Pending cases are treated as unresolved; no outcomes are predicted.

## II. The Legal Frameworks

### A. United States

Copyright (established law). Model training involves copying, which falls within the exclusive reproduction right in 17 U.S.C. § 106(1). The central question is therefore often whether the copying qualifies as fair use under § 107.[9]

In *Bartz v. Anthropic*, Judge Alsup described training on books as highly transformative and found it to be fair use. He also found that digitizing print books that had been lawfully purchased was fair use. But he did not consider the later transformative purpose enough to excuse creating a permanent central library from pirated copies.[10]

In *Kadrey v. Meta*, Judge Chhabria also ruled for the defendant, but emphasized that the record was limited. He cautioned against giving transformativeness too much weight and treated market harm as the decisive factor. He also discussed possible markets plaintiffs might rely on, including a theory that AI training could dilute existing markets.[11]

The result in *Thomson Reuters v. ROSS* was different. That case involved a non-generative legal-search tool. The district court found no fair use because ROSS had built a competing product.[12] On September 29, 2026, the Third Circuit affirmed, finding that three fair-use factors weighed against ROSS. It also accepted that damage to a developing market for

licensing headnotes as AI training data could count as market harm.[13] ROSS has said it plans to seek Supreme Court review.

The U.S. Copyright Office's pre-publication Part 3 report, released May 9, 2025, does not endorse a categorical rule. It treats market dilution as a possible form of harm and notes that copying entire works weighs against fair use. The report is persuasive rather than binding authority, was released under unusual circumstances, and is not final.[14]

Privacy (established law). The United States has no single, comprehensive federal privacy statute. Privacy limits on AI training instead arise from state laws, the Federal Trade Commission's authority under Section 5, and sector-specific rules.

The California Consumer Privacy Act excludes "publicly available information" from its definition of personal information,[15] limiting its reach over some scraping. Illinois's Biometric Information Privacy Act requires notice and written consent before collecting biometric identifiers, making it a significant statutory constraint on datasets built through face scraping.[16] The FTC has also required companies to delete models and algorithms derived from improperly obtained data. Such remedies, sometimes called "algorithmic disgorgement," have appeared in consent orders such as *Everalbum*.[17]

Interpretation. In practice, copyright litigation is the main legal constraint on training in the United States. Privacy rules play a more limited role, mainly through state statutes, some of which carve out publicly available data, and through FTC enforcement remedies.

**B. European Union**

Copyright (established law). Directive 2019/790 creates two TDM exceptions. Article 3 applies to research organizations and cultural heritage institutions and cannot be overridden by contract. Article 4 is available to other users with lawful access, unless the rights holder has reserved the relevant rights in an appropriate way. For online material, this may include a machine-readable reservation.[18]

In *Kneschke v. LAION*, the Hamburg Higher Regional Court dismissed a photographer's appeal on December 10, 2025. It found that creating the dataset fell within the TDM exceptions and that the reservation on the source website was ineffective. Commentators disagree about the court's reasoning on the form of the reservation and whether it needed to be machine-

readable.[19] A further appeal to the Bundesgerichtshof is pending. The court heard it on September 3, 2026, and one secondary source reports that a decision is expected on December 17, 2026.[20]

In *GEMA v. OpenAI*, the Munich Regional Court found on November 11, 2025, that lyrics memorized by a model were reproduced in the model's parameters and again in its outputs. It interpreted the TDM exception narrowly. OpenAI has appealed.[21]

The Court of Justice of the European Union is considering *Like Company v. Google Ireland* (C-250/25), described as the first referral concerning generative AI and copyright. The questions include whether training and chatbot outputs involve reproduction and whether Article 4 applies. No decision had been found for this paper.[22]

The AI Act requires general-purpose AI providers to maintain a copyright-compliance policy that identifies and respects reservations made under Article 4(3). It also requires them to publish a sufficiently detailed summary of the content used to train their models.[23]

Privacy (established law). Under the GDPR, information remains personal data if it relates to an identifiable person, even if it is publicly available.[24] Training on scraped personal data therefore requires a lawful basis—often argued to be legitimate interest under Article 6(1)(f)—and compliance with requirements such as fairness, transparency, data minimization, and purpose limitation. Special-category data requires a separate condition under Article 9. Controllers may also have to provide information to individuals under Article 14 and respect rights to erasure and objection under Articles 17 and 21.

In Opinion 28/2024, the EDPB concluded that legitimate interest may, in principle, support AI development and deployment if a three-part assessment is satisfied. It also said that whether a model is anonymous must be assessed case by case. A model is anonymous only where it is very unlikely both that individuals can be identified directly or indirectly and that personal data can be extracted through queries. The EDPB further stated that unlawful processing during development may affect the lawfulness of later deployment unless the model has been properly anonymized.[25]

The CJEU has held that a purely commercial interest can qualify as a legitimate interest under Article 6(1)(f).[26] On November 19, 2025, the Commission proposed a Digital Omnibus that

would expressly recognize legitimate interest as a possible basis for AI training. As of late September 2026, the data-law provisions had not been adopted, and negotiations in the Council were continuing. Reports differed on whether the explicit provision would remain in the final text.[27]

**C. India**

Privacy (established law, partly in force). The Supreme Court recognized privacy as a fundamental right in *Puttaswamy*.[28] The Digital Personal Data Protection Act, 2023 is centred on consent. Section 6 requires consent to be free, specific, and informed, and allows it to be withdrawn as easily as it was given. Section 7 lists certain "legitimate uses," but does not establish a general commercial legitimate-interest ground.[29]

Section 3(c)(ii) excludes personal data that the individual has made publicly available, or that another person has published because the law requires it.[29] The DPDP Rules, 2025 were notified in November 2025 and set out a phased start date: provisions concerning the Data Protection Board took effect immediately, consent-manager provisions follow after twelve months, and most substantive duties begin after eighteen months—around May 13, 2027.[30]

Interpretation. The Act does not explain whether using personal data for AI training changes the application of the Section 3(c)(ii) exclusion. Since the main duties are also not yet enforceable, the law's treatment of AI training remains uncertain.

Copyright (established law, evolving). Under Section 51 of the Copyright Act, 1957, unauthorized reproduction infringes copyright unless an exception in Section 52 applies. Section 52(1)(a) allows fair dealing for listed purposes, including private use or research, criticism, and review. The Act contains no TDM exception.[31]

In *ANI Media v. OpenAI*, a single judge of the Delhi High Court refused interim relief on July 24, 2026. The judge took the preliminary view that training could fall within fair dealing, while cautioning that this did not mean a licence would never be required. ANI appealed, and a Division Bench issued notice on September 15, 2026.[32] The order itself was not available for this article, so the account relies on secondary reports.

In December 2025, a Department for Promotion of Industry and Internal Trade committee published the first part of a working paper. It recommended a "hybrid model": a mandatory

blanket licence, with statutory remuneration for all lawfully accessed works used in training and no opt-out right. This remains a proposal, not law.[33]

## III. How the Regimes Interact

The table summarizes the permission routes that may apply to the same dataset.

| Jurisdiction | Copyright route | Privacy route | Overall effect |
|---|---|---|---|
| United States | Fair use, assessed case by case; courts have reached different conclusions, and the Third Circuit has recognized potential harm to licensing markets | Few federal privacy limits; state laws with public-data exclusions, BIPA, and FTC remedies | Copyright is the main constraint |
| European Union | Articles 3 and 4 TDM exceptions, including an opt-out; AI Act policy and summary duties | GDPR lawful-basis requirement, Article 9 conditions, Article 14 transparency, and individual rights | Two separate compliance tracks apply |
| India | Fair dealing under Section 52(1)(a); proposed blanket licence | DPDP consent rules, limited Section 7 uses, Section 3(c)(ii) exclusion, and duties phased in through 2027 | Significant gaps and unresolved questions |

### A. Provenance: one fault line, different meanings of "lawful"

Established law. *Bartz* held that the transformative character of later use did not excuse piracy of the source copies.[10] The EU's TDM exceptions require lawful access.[18] The DPIIT proposal would cover only "lawfully accessed" works.[33] The EDPB has also said that unlawful processing during a model's development can affect the lawfulness of its later deployment.[25]

Interpretation. Each system treats the way data was acquired as an important threshold, but "lawful" means something different in copyright and privacy law. For copyright, it generally concerns authorization or whether access involved pirated copies. For privacy, it concerns having a lawful basis and meeting requirements such as transparency and purpose limitation.

A dataset can pass one test and fail the other. For example, a news archive might be lawfully accessed for TDM but contain sensitive details about private individuals without an established basis under GDPR Articles 6 or 9. Conversely, developers might have valid consent to process data uploaded by users, while those uploads include photographs belonging to someone else. A provenance audit for one area of law cannot be assumed to satisfy the other.

**B. Opt-outs and the wrong rights holder**

The EU copyright opt-out belongs to the rights holder. It allows that person to reserve rights in an appropriate way, including through machine-readable signals for online content.[18] The GDPR rights that may appear similar belong to data subjects: individuals can object to some processing under Article 21 and request erasure under Article 17.[24]

Interpretation. These rights belong to different people, are exercised through different channels, and do not have the same effect. A photographer's reservation on a website does not necessarily protect a person shown in the photograph. Likewise, a data subject's objection does not automatically bind a scraper through a crawler directive, since the website operator may be neither the author nor the person whose data is being processed.

The Hamburg case also illustrates the uncertainty that can surround a formal reservation, even within copyright law. The result depended on how the reservation was expressed, and the court's approach has been described differently by commentators.[19] Copyright reservation mechanisms therefore cannot simply be reused for privacy objections. Each requires its own process and accountability.

India presents a related tension. The DPIIT proposal would eliminate an opt-out for copyright holders,[33] while the DPDP Act allows consent to be withdrawn.[29] Interpretation: if a statutory licence permits a work to be used for training, but that work contains personal data of someone who later withdraws consent, the two approaches may point in different directions. The proposal does not explain which would prevail.

**C. Memorization: how both regimes may reach the model**

Established law. In *GEMA*, the court treated memorized lyrics as reproduced in the model and again in its outputs.[21] The EDPB, meanwhile, rejects a general assumption that models trained on personal data are anonymous, and calls for an assessment of whether personal

information can be extracted through queries.[25] *Bartz* and *Kadrey* focused on training as a use and treated substitution through outputs as a separate issue, though *Kadrey* also discussed market dilution.[10][11]

Interpretation. Retaining and reproducing training material is where both areas of law may treat a model as a regulated object, rather than looking only at the training process. Their remedies differ. Copyright law may lead to injunctions or damages. Data-protection law may require erasure, restricted processing, retraining, or deletion of a model that is not anonymous. U.S. consent orders may also require algorithmic disgorgement.[17]

A developer's effort to reduce memorization may help with both legal concerns, but the tests are different. Copyright asks whether a work has been reproduced or whether there is substantial similarity. Privacy law asks whether a person can be identified and their data extracted. Shared testing could avoid duplicating some work, but no common regulatory protocol is identified here.

**D. Licensing, public availability, and the risk to data subjects**

The Third Circuit's recognition of a developing licensing market for AI training data strengthens the argument for paid copyright licences.[13] Interpretation: this may benefit rights holders, but it does not settle the privacy issues. A copyright licence conveys the licensor's copyright interests. It does not necessarily establish a lawful basis for personal data in the licensed material, satisfy Article 14 transparency requirements, or bind data subjects who are not parties to the agreement. If a licence comes from a data broker or platform, the developer may receive data without knowing what basis supports its processing.

The effect of public availability varies across the three jurisdictions. In the EU, public availability does not take data outside the GDPR, so lawful access under TDM rules and lawful processing under the GDPR remain separate questions. In the United States, public-data exclusions in some state privacy laws can reduce those laws' application,[15] leaving copyright and FTC remedies as the more effective constraints.

In India, Section 3(c)(ii) excludes some data that individuals publish themselves,[29] but the same user-generated content may remain copyright-protected and fall within the DPIIT proposal. The position differs for personal data published by a third party: the Act's exclusion

would not apply, and the Act does not provide a general legitimate-interest basis, leaving consent or a specified Section 7 use as possible routes.[29]

Interpretation. Read literally, the Indian framework could be relatively permissive for data individuals publish themselves while being more restrictive for data published by others. That distinction may be difficult to apply to large-scale scraping. Secondary commentary also notes that the Act does not resolve how aggregation, inference, or AI development affects the Section 3(c)(ii) exclusion.[34] No authority resolving that question was identified.

## IV. Counterarguments and Responses

1. "The regimes are separate, so there is no benefit in studying their interaction." The strongest version of this argument is that copyright protects expression while privacy law protects people, allowing each field to be applied independently.

Response. The doctrines differ, but their practical operation overlaps. Both apply to the same dataset and model weights. They also raise related questions about provenance and documentation: the AI Act's training summary, GDPR Article 14 disclosures, and copyright access records may all concern the same corpus. Their remedies can also conflict. The fact that the rules are legally distinct is a reason to assess them separately and cumulatively, not to ignore their interaction.

2. "Models learn statistical patterns rather than storing works or personal data, so only the initial copying matters." This view is consistent with the emphasis on transformative use in *Bartz* and *Kadrey*.

Response. Whether a model retains or reveals its training data is an empirical question that can vary by model, dataset, and prompt. *GEMA* found memorization on the record before it, and the EDPB says anonymity must be demonstrated rather than assumed.[21][25] Even if typical outputs do not reproduce training material, both legal regimes can still apply to the training copy itself: copyright law addresses reproduction in the United States and EU, and the GDPR addresses processing.

3. "Strong privacy enforcement will slow innovation and entrench large companies; the law should expressly recognize legitimate interest." The Commission's Omnibus proposal and remarks by the Delhi court about the practical difficulty of obtaining licences from many

parties reflect this concern.[27][32]

Response. This is a serious policy argument, and current EDPB guidance already allows legitimate interest in principle. But codifying that possibility while retaining case-by-case balancing and data-subject rights would shift the analysis rather than remove it.[25][27] Clearer rules—such as standard assessment templates, dataset documentation, and regulatory guidance—could give smaller developers more predictability without creating a presumption that training is lawful.

4. "A mandatory blanket licence would simplify access and pay creators, solving both problems."

Response. A blanket licence might address some copyright transaction costs, as the DPIIT paper proposes. Its effectiveness would still depend on a workable system for distributing remuneration, which critics question.[33] It would not address the rights of people whose data appears in licensed works. Removing an opt-out could also heighten tension with the DPDP Act's right to withdraw consent, as discussed in Part III.B.

## V. Recommendations and Areas for Further Research

The following are recommendations, not statements of established law.

1. Keep one provenance record and conduct two legal assessments. Developers could maintain a shared dataset record to support AI Act summaries, GDPR disclosures and processing records, and copyright access logs. They should still assess privacy and copyright compliance separately.

2. Do not treat a licence as privacy compliance. Contracts should allocate responsibility for personal data in a licensed corpus, including warranties about the lawful basis, data minimization, and cooperation with deletion requests. Licensees should also conduct their own lawful-basis assessment.

3. Draft legislation to preserve individual rights. If the EU expressly recognizes legitimate interest for AI training, legislation should retain the balancing test, objection rights, and transparency duties. If India adopts a blanket licence, it should state that the licence does not displace the DPDP Act and clarify how Section 3(c)(ii) applies to training datasets before the Act's main duties begin in 2027.

4. Coordinate regulatory guidance. The AI Office and data-protection authorities should jointly address memorization and extraction testing. A shared technical protocol could help inform both copyright infringement and privacy anonymity assessments.

5. Create separate processes for copyright reservations and privacy objections. Machine-readable signals can communicate rights holders' reservations. Data subjects need a separate and accountable way to exercise their rights.

Further research. Useful next steps include: (a) measuring memorization rates across model types and relating the results to legal tests; (b) assessing how the Third Circuit's *ROSS* decision and its licensing-market analysis affect generative AI cases, including any Supreme Court review; (c) examining the CJEU's decision in *Like Company* and the BGH's ruling in *Kneschke*; (d) tracking how the Delhi Division Bench addresses *ANI* and Section 3(c)(ii), especially once the DPDP Act's main duties take effect; and (e) studying whether courts apply consistent standards when deciding if a model is a "copy" or contains "personal data."

## VI. Conclusion

The 2025–2026 decisions point to several recurring themes: the source of training data matters; licensing markets may be relevant to market harm; and memorization can make a model itself the subject of legal scrutiny. But the decisions have not settled how copyright and data-protection obligations fit together.

For developers, the central practical point is that the two areas require separate compliance assessments built around a shared record of the training data. For legislators, copyright licensing answers questions about works; it does not resolve questions about individuals. Proposals such as India's mandatory blanket licence and the EU's possible recognition of legitimate interest should make that distinction clear. The next stage of the law will depend in part on the pending matters in *Like Company*, *Kneschke*, and *ANI*, as well as any Supreme Court review of *ROSS*.

## VII. Assumptions and Limitations

1. Research method. The author reports checking authorities through web searches on October 28, 2026. For most court decisions, the primary judgment was not available; the account relies on law-firm analysis, case trackers, and institutional pages. Pinpoint

references drawn from secondary sources should be checked against the judgments or reporters before submission.

2. Authorities that are not final. The *ROSS* decision may receive further review; *Kneschke* remains pending before the BGH; *GEMA* is under appeal; *Like Company* remains before the CJEU; *ANI* concerns an interim order under appeal; the EU Data Omnibus has not been adopted; the DPIIT paper is only a recommendation; and the Copyright Office Part 3 report is a pre-publication report. None should be presented as settled law.

3. Scholarship. Although the research prompt called for reputable academic sources, the author did not include peer-reviewed scholarship by author because the relevant authors, titles, and pinpoint citations could not be verified. The bibliography contains leads to confirm. The analysis relies primarily on legal authorities and official guidance.

4. Scope. The article addresses training and memorization. It does not examine output liability in depth, authorship of AI outputs, biometric rules beyond a brief reference, or trade-secret and contract restrictions on scraping.

5. Interpretation. Sections labelled “Interpretation” and the recommendations in Part V reflect the author’s analysis, not settled law.

**Endnotes:**

[1] *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1022 (N.D. Cal. 2025); *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1051 (N.D. Cal. 2025).
[2] *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc.*, No. 25-2153 (3d Cir. Sept. 29, 2026) (affirming partial summary judgment).
[3] Oberlandesgericht Hamburg [OLG] [Higher Regional Court] Dec. 10, 2025, 5 U 104/24 (Ger.), aff'g Landgericht Hamburg [LG] [Regional Court] Sept. 27, 2024, 310 O 227/23 (Ger.).
[4] Landgericht München I [LG] [Munich Regional Court] Nov. 11, 2025, *GEMA v. OpenAI* (Ger.) (docket number not verified).
[5] Eur. Data Prot. Bd., Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models ¶ 34 (adopted Dec. 2024).
[6] Regulation 2024/1689, recital 106, 2024 O.J. (L 2024/1689) [hereinafter AI Act].
[7] Regulation 2016/679, art. 3(2), 2016 O.J. (L 119) 1 [hereinafter GDPR].
[8] Digital Personal Data Protection Act, 2023, No. 22 of 2023, § 3(b) (India) [hereinafter DPDP Act].
[9] 17 U.S.C. §§ 106(1), 107.
[10] *Bartz*, 787 F. Supp. 3d at 1022, 1025.
[11] *Kadrey*, 788 F. Supp. 3d at 1051.
[12] *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc.*, No. 1:20-cv-00613 (D. Del. Feb. 11, 2025) (Bibas, J., sitting by designation) (reporter citation not verified).
[13] See supra note 2.
[14] U.S. Copyright Off., *Copyright and Artificial Intelligence, Part 3: Generative AI Training* (pre-publication version May 9, 2025).
[15] Cal. Civ. Code § 1798.140(v)(2) (West) (subsection number from memory; not verified this session).
[16] 740 Ill. Comp. Stat. 14/15(b) (from memory; not verified this session).
[17] 15 U.S.C. § 45; *In re Everalbum, Inc.*, FTC File No. 192-3172 (May 6, 2021) (decision and order) (docket and date from memory; not verified this session).
[18] Directive 2019/790, arts. 3, 4(1)–(3), 2019 O.J. (L 130) 92.
[19] OLG Hamburg, 5 U 104/24 (Ger.). Secondary sources describe the court's reasoning on the reservation differently; consult the judgment.
[20] Bundesgerichtshof [BGH] [Federal Court of Justice], I ZR 281/25 (Ger.) (pending;

hearing date and judgment date reported by a single secondary source).

[21] *GEMA v. OpenAI*, supra note 4 (appeal to the Munich Higher Regional Court reported pending).

[22] Case C-250/25, *Like Company v. Google Ireland Ltd.* (CJEU) (referral of Apr. 3, 2025 by the Budapest Környéki Törvényszék) (pending).

[23] AI Act art. 53(1)(c)–(d).

[24] GDPR arts. 4(1), 5, 6(1)(f), 9, 14, 17, 21.

[25] EDPB Opinion 28/2024, supra note 5; Eur. Data Prot. Bd., *EDPB Opinion on AI Models: GDPR Principles Support Responsible AI* (press release, Dec. 18, 2024).

[26] Case C-621/22, *Koninklijke Nederlandse Lawn Tennisbond v. Autoriteit Persoonsgegevens* (CJEU Oct. 4, 2024) (from memory; not verified this session).

[27] European Commission, *Digital Omnibus* (proposal published Nov. 19, 2025) (document number not verified). For status, see Heuking, *GDPR, AI, and Cookies: What the EU's Data Omnibus Could Change for Companies* (Sept. 24, 2026); Noerr, *Digital Omnibus: AI Is Nearing the Finish Line* (June 18, 2026); IAPP, *A View from Brussels: Simplification?* (secondary).

[28] *K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1 (India) (from memory; no pinpoint).

[29] DPDP Act §§ 3(c)(ii), 6, 7.

[30] Digital Personal Data Protection Rules, 2025 (India) (notified Nov. 13–14, 2025; sources differ on the date).

[31] Copyright Act, 1957, No. 14 of 1957, §§ 51, 52(1)(a) (India).

[32] *ANI Media Pvt. Ltd. v. OpenAI Inc.*, CS(COMM) 1028/2024 (Del. H.C. July 24, 2026) (interim order) (India) (known only through secondary reports; the order was not retrieved).

[33] Dep't for Promotion of Indus. & Internal Trade, *Working Paper on Generative AI and Copyright, Part 1: One Nation One Licence One Payment* § 5.1 (Dec. 2025) (India).

[34] India Briefing, *India DPDP Act and Publicly Available Personal Data* (commentary; secondary source).