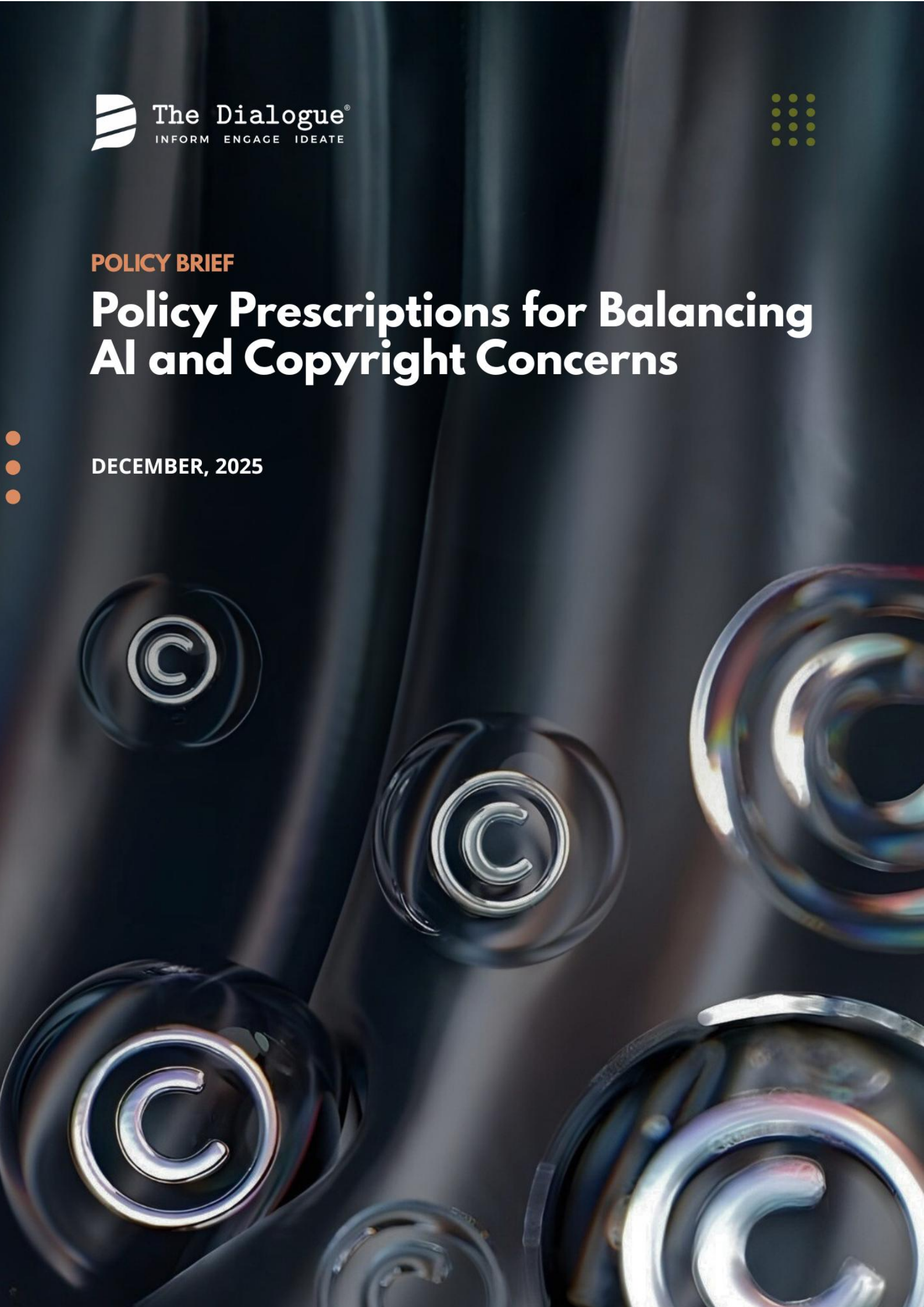
The Dialogue®
INFORM ENGAGE IDEATE

POLICY BRIEF

Policy Prescriptions for Balancing AI and Copyright Concerns

DECEMBER, 2025



POLICY BRIEF

POLICY PRESCRIPTIONS FOR BALANCING AI AND COPYRIGHT CONCERNS

Author: Jameela Sahiba

Thematic Designer : Shivam Kulshrestha

The Dialogue® is a public policy think tank with a vision to drive a progressive narrative in India's policy discourse. Founded in 2017, we believe in facilitating well-researched policy debates at various levels to help develop a more informed citizenry, on areas around technology and development issues. The Dialogue® has been ranked as the world's Top 10 think tanks to watch out for, by the Think Tank and Civil Societies Programme (TTCSP), University of Pennsylvania in their 2020 and 2021 rankings.

For more information
www.thedialogue.co

Suggested Citation
Sahiba, J. (December, 2025). Policy Brief: Policy Prescriptions for Balancing AI and Copyright Concerns. The Dialogue.®

Catalogue No
TD/ET/PB/1225/15

Publication Date
December 17, 2025

Context

I. Introduction

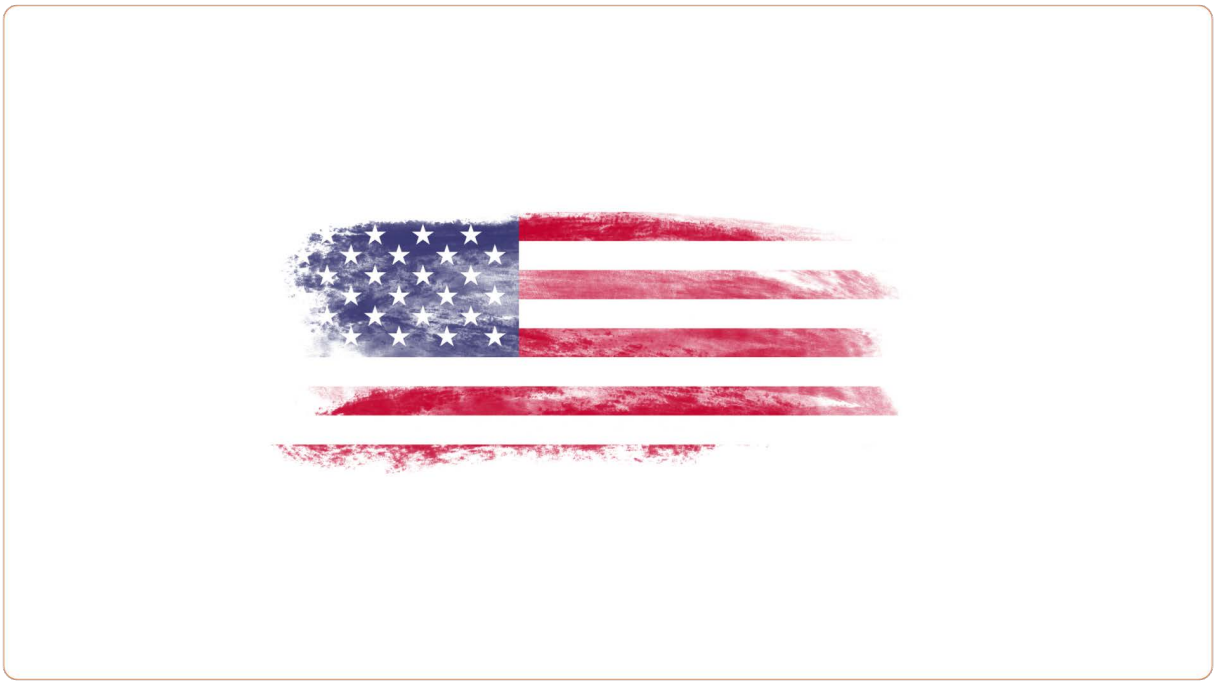
The crossroads between copyright and artificial intelligence (hereinafter referred as 'AI') innovation is emerging as one of the most complex and contested domains in contemporary digital policy. The acceleration of the information revolution, manifested in the staggering daily production of tweets, emails, social media content, instant messages, and audiovisual uploads, has created unprecedented bodies of data that can be used to teach and improve AI systems. This ever-expanding corpus of human-generated knowledge presents both opportunity and tension. While AI models require massive datasets, often composed of copyright-protected works, for training and refinement, the use of such material invites deep legal and ethical scrutiny on how to balance the legitimate interests of creators with the needs of innovators developing next-generation AI applications.

Drawing from global approaches, the analysis includes the European Union's text and data mining opt-out regime, the United Kingdom's fair dealing exceptions, Japan's data mining laws, the United States' fair use doctrine, China's moderate leniency approach, to name a few. The comparative review provides insight into the strengths and pitfalls of each model, offering critical lessons for their potential adaptation within India's socio-legal context.

The objective is to inform Indian policymakers on approaches that protect creative rights without stifling technological advancement, and to identify pathways that allow creators to participate more fully in emerging digital markets. This inquiry culminates in the development of actionable policy prescriptions for three primary stakeholder groups: policymakers, AI developers, and rights holders. Through a careful review of global models and India-specific needs, the paper seeks to guide the formulation of copyright policy that supports innovation while upholding the principles of fairness and cultural integrity.

II. Jurisdictional Analysis

For each jurisdiction, this paper summarizes (1) the legal instrument, (2) what it means for AI training, (3) the principal opportunities (pros) that approach creates for AI developers, rightsholders and the public, (4) the main challenges (cons) and legal risk vectors and (5) decisive or illustrative court decisions (if any) or developments that are already shaping outcomes.



1. United States: fair use (Section 107) and a case-by-case model

1.1 Legal Instrument

In the United States, copyright law employs a flexible, open-ended framework for addressing unauthorized uses of protected works through the doctrine of "fair use" codified at S. 107 of the Copyright Law of the United States[1]. This doctrine empowers courts to evaluate potentially infringing acts on a case-by-case basis by weighing four statutory factors: (1) the purpose and character of the use, particularly whether the use is transformative or commercial; (2) the nature of the copyrighted work, distinguishing between factual and creative content; (3) the amount and substantiality of the portion used relative to the whole work; and (4) the effect of the use on the potential market for or value of the original work. This balancing test provides a mechanism for courts to excuse certain secondary uses without prior authorization, thereby encouraging socially valuable activities such as teaching, criticism, commentary, and research. Applying the fair use test, courts have held lawful new technologies such as search

[1] Copyright Law of the United States (Title 17) S. 107 <https://www.copyright.gov/title17/92chap1.html#107> last accessed 9th october 2025

engines,[2] books digitization for students with print-disabilities,[3] and plagiarism detection software.[4]

1.2 Implications for AI training

The practical significance of fair use in the context of text and data mining (hereinafter referred to as 'TDM') and GenAI training lies in the fact that the U.S. has not adopted a statutory TDM exception comparable to those in some other jurisdictions. Instead, AI model developers primarily rely on fair use as a defense when using copyrighted texts, images, and other materials to train their systems. This reliance subjects AI training to judicial scrutiny on a fact-specific basis, with the permissibility of training AI on copyrighted content depending largely on whether courts find the use sufficiently transformative and non-substitutive of the original works. While the fair use framework grants significant leeway for innovation, it entails legal uncertainty, as outcomes turn on nuanced fact patterns and judicial interpretations that may evolve with the increasing AI litigation volume.

1.3 Opportunities / Pros

- A key benefit of the U.S. fair use framework is that it can accommodate new technologies that interact with copyrighted works without new legislation; courts can find those technologies lawful because their use of copyrighted works is fair. Through fair use opinions, courts can craft case-specific rulings that reflect the unique considerations presented by the technology. Judicial precedents, such as *Google/HathiTrust*[5], demonstrate that courts are willing to recognize transformative uses of copyrighted works that serve public interest and advance knowledge. This innovation-friendly stance enables AI developers to experiment with novel applications and architectures without awaiting legislative updates, while still operating within the guardrails of copyright law. This is true for both AI training and AI applications that use copyrighted works which means fair use is flexible enough to provide guidance on both.
- In the context of AI training, the flexibility of fair use supports robust commercial and research applications of AI while developers must keep in mind careful compliance with copyright norms and potential evidentiary burdens if called to defend their practices. This allows commercial actors to operate without negotiating licenses in many scenarios, lowering transaction costs and accelerating R&D.
- Since no prior authorization or machine-readable opt-out regime is required under federal law, public and scraped datasets may be used, subject to the fair use analysis.

---

[2] *Perfect 10 v. Amazon (2011)* <https://law.justia.com/cases/federal/appellate-courts/ca9/10-56316/10-56316-2011-08-03.html>

[3] *Authors Guild v. HathiTrust (2014)* <https://law.justia.com/cases/federal/appellate-courts/ca2/13-4829/13-4829-2015-10-16.html?> last accessed on 9th October, 2025

[4] *A. V. v. iParadigms, LLC (2009)* <https://law.justia.com/cases/federal/appellate-courts/ca4/08-1424/081424.p-2011-03-15.html>

[5] *Authors Guild v. HathiTrust (2014)* <https://law.justia.com/cases/federal/appellate-courts/ca2/13-4829/13-4829-2015-10-16.html?> last accessed on 9th October, 2025.

1.4 Challenges / Cons

- Fair use is inherently fact-driven. Businesses face litigation risk and unpredictable outcomes; decisions can differ by judge or circuit.
- Even though the law ultimately favours defendants, discovery battles and litigation expenses are substantial.
- Recent rulings indicate courts may scrutinize how data was acquired.

1.5 Key caselaw shaping the field

- **Authors Guild v. Google / HathiTrust (2nd Cir., 2015 / 2014)[6]:** This case provides important precedents finding large-scale digitization and non-display search uses to be fair use because they were transformative and did not supplant the markets for the original works. These remain the most analogous precedents supporting TDM claims of fair use.
- **Bartz v. Anthropic (N.D. Cal., June 2025)[7]:** Judge granted summary judgment for Anthropic as to lawful purchases that were scanned (*a format-shift*) and as to training generally on the ground that training is transformative; denied summary judgment for use of pirated copies and left those issues for trial. The opinion recognizes a fair-use path for model training but still leaves open questions related to source acquisition.
- **Kadrey v. Meta (N.D. Cal., June 25, 2025)[8]:** In this partial summary judgment for Meta, the court found that the plaintiffs failed to show market harm and that training could be transformative, though factual disputes remain, and plaintiffs may still pursue limited claims. The ruling echoes the Anthropic analysis.

1.6 Practical takeaway from U.S.

Fair use permits an expansive, innovation-friendly approach but imposes litigation risk. Firms operating in the U.S. can often proceed without prior licensing, but many may choose to negotiate licenses, implement dataset hygiene policies (reject obvious pirated sources), or obtain indemnities to reduce risk.

[6] *Ibid*

[7] *Bartz v. Anthropic PBC* (2025) <https://www.jw.com/news/insights-kadrey-meta-bartz-anthropic-ai-copyright/> last accessed on 9th October, 2025

[8] *Kadrey v. Meta Platforms, Inc.*(2025) <https://www.jw.com/news/insights-kadrey-meta-bartz-anthropic-ai-copyright/> last accessed on 9th October, 2025

2. European Union: statutory TDM exception (Directive 2019/790) + AI Act overlay

2.1 What the law is

The EU's 2019 Copyright in the Digital Single Market (DSM) Directive[9] creates an exception for TDM when done by research organizations and cultural heritage institutions and a broader exception in Article 4 that allows TDM of lawfully accessible content for any purpose, subject to the ability of rightholders to "expressly reserve" rights against text and data mining in an "appropriate manner" (defined as machine-readable means for works available on the internet) . The DSM regime is therefore a statutory, rule-based approach rather than a judge-driven balancing test. Each country in the European Union has implemented these two exceptions into their national copyright laws.

The EU AI Act (2024)[10] further adds a regulatory layer, ie., it creates a separate regulatory obligation for providers of general-purpose AI (GPAI) models. Under the AI Act, GPAI model providers are required to "put in place a policy to comply with Union law on copyright," including the TDM exception in the DSM Directive and the reservation of rights. To help companies comply with this requirement, the European Commission in July 2025 published the Code of Practice (CoP)[11] for General-Purpose AI Models. The CoP outlines a series of steps that GPAI providers can take to meet the AI Act's legal obligations..

[9] Digital Single Market and amending Directives 96/9/EC and 2001/29/EC <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?> last accessed on 9th October, 2025

[10] European Union's Artificial Intelligence Act, Regulation (EU) 2024/1689 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689> last accessed on 9th October, 2025

[11] General-Purpose AI Code of Practice (2025) <https://digital-strategy.ec.europa.eu/en/news/general-purpose-ai-code-practice-now-available> last accessed on 9th October, 2025

2.2 Implications for AI training

Commercial TDM is allowed in principle under *Article 4* but rightholders can expressly reserve their rights (i.e., opt out of text and data mining). But the DSM Directive does not explain what are "appropriate means" of reserving rights, and there has been little litigation on the question, so significant legal uncertainty remains about what works can legally be used for AI training. Rightsholders and AI developers have very different views on what forms of opt-out qualify under the exception or are technically feasible. Additionally, some rightsholders have begun to argue that AI training does not qualify as "text and data mining," which would mean that AI training on copyrighted works is not permitted in the EU without authorization from rightsholders. Additionally, because the EU AI Act incorporates copyright obligations, developers need to think about legal compliance both in terms of copyright litigation in the courts of EU member states, as well as in regulatory enforcement at the EU level by the European Commission.

2.3 Opportunities / Pros

- A statutory exception theoretically gives clearer rules about when TDM is permitted compared with ad hoc fair use balancing.
- The opt-out mechanism allows creators and data owners to preserve licensing revenues if they want to, which offers an attractive feature for cultural industries.
- Because many works are protected by copyright but their owners do not intend to commercially exploit them, those works are unlikely to be opted out of training, and therefore will be available for AI training. This provides data that AI developers can use without copyright risks.

2.4 Challenges / Cons

- There is deep disagreement between AI developers and rightsholders about how Article 4 opt-out works, such as A) whether it must be explicitly invoked or is exercised by legal statements reserving "all rights," B) whether it can be exercised by collective management organizations on behalf of their members, and C) whether the reservation needs to be connected to specific works, or whether it is sufficient rightsholder to state they are reserving rights for their repertoire, leaving AI developers to determine what works that corresponds with. These questions are not answered by the Directive or EU courts, which makes dataset assembly challenging Critics[12] argue the opt-out undermines the commercial TDM exception's utility.
- Opt-out systems give the final decision over what data is used for AI training to rightsholders. If a rightsholder owns particularly valuable data (such as scientific publications), or acquired the rights to a large body of data in a particular domain,

[12] Artha Dermawan, 'Text and Data Mining Exceptions in the Development of Generative AI Models: What the EU Member States could Learn from the Japanese "Nonenjoyment" Purposes?' (2023) 27 The Journal of World Intellectual Property 44, 53. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jwip.12285> last accessed on 9th October, 2025

they could demand extraordinary licensing payments or refuse to license that data and prevent AI models from gaining knowledge about a specific field.

- An opt-out system may entrench the power of large technology companies. Large platforms can negotiate licenses more easily than smaller players, potentially entrenching incumbents. Large platforms also have the resources to build custom tooling to detect opt-outs, depending on how complex opt-out systems are to operate. Finally, large platforms may not need to license as much third-party data: social media companies, for example, have terms of service that allow them to build and improve products with data from their users.
- There are few machine-readable forms of rights reservation today. The main standard used by both rightsholders and AI developers is the website protocol robots.txt, which some rightsholders have criticized because the website owner who controls that protocol may not own the rights to all of the content on their website.

2.5 Key caselaws/developments shaping the field

- The Hamburg District Court ruled in the *Kneschke v. LAION*[13] case that LAION's web scraping of publicly available images for creating an AI training dataset did not infringe copyright. The court deemed this activity a legitimate form of TDM for scientific research, protected under Germany's implementation of the DSM Directive's Article 3 TDM exception (Section 60d of the UrhG[14]). It rejected the plaintiff's argument that web scraping for AI is inherently commercial and thus subject to an opt-out, stating that copyright analysis must distinguish between the dataset creation, model training, and the subsequent application of the AI. This case is currently on appeal.
- The *Like Company v. Google Ireland*[15] case raises many important questions, including whether training LLMs constitutes an action reproduction under EU copyright law, whether AI training constitutes a form of text and data mining, and if AI-generated answers using internet content constitutes a communication to the public. These questions were referred to the European Court of Justice. A ruling favoring the Like Company could mandate extensive licensing for AI development, while a ruling for Google could confirm that the act of AI training does not implicate a copyright owner's right of reproduction. The decision is expected to be delivered before 2027.

2.6 Practical takeaway from EU

The EU's approach privileges rule clarity and rightholder control over flexibility for AI research and experimentation. By reserving rights, rightsholders can block any form of TDM that is not conducted by a research organization or cultural heritage institution. However, the EU approach also shows the dangers of failing to think through how opt-outs work in practice; European rightsholders currently believe a wide variety of opt-out methods are valid but being

[13] LAION v Robert Kneschke, [27 September 24] <https://www.euipo.europa.eu/en/law/recent-case-law/germany-hamburg-district-court-310-o-22723-laion-v-robert-kneschke> last accessed on 9th October, 2025

[14] Act on Copyright and Related Rights (Urheberrechtsgesetz – UrhG) <https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html> last accessed on 9th October, 2025

[15] Like Company v. Google Ireland (Case C-250/25) <https://www.lexology.com/library/detail.aspx?g=98251adb-24e0-4977-948a-04ab185fb138> last accessed on 9th October, 2025

ignored, and AI companies believe that many opt-out attempts are invalid but could still be the basis for copyright litigation or regulatory enforcement.

3. United Kingdom: narrow, research-only TDM exception and policy flux

3.1 What the law is

The UK does not have a copyright exception for commercial text and data mining. The UK formally began withdrawing from the EU in 2017, so it never implemented the 2019 DSM Directive, including the broad exception for commercial TDM under Article 4. Instead, the UK's Copyright, Designs and Patents Act contains a TDM/"computational analysis" exception (section 29A)[16] that was enacted in 2014 and is limited to non-commercial research. . Seeking to spur AI development, the government opened a public consultation in late 2024 on whether to add a new copyright exception for commercial text and data mining. The government suggested different possible approaches to a TDM exception and announced a preference for an exception with an opt-out right, similar to the EU. Rightsholders strongly objected to *any* new TDM exception, including one with an opt-out, and mounted a public campaign against the proposal. The government is still reviewing the 11,000 public submissions to the consultation, but in the meantime it is planning technical discussions between rightsholders and AI developers as the next step of the process.

[16] Copyright, Designs and Patents Act 1988 <https://www.legislation.gov.uk/ukpga/1988/48/section/29A> last accessed on 9th October, 2025

While this consultation is pending, Parliament has started to get involved. In June 2025, there was a series of amendments to the *Data (Use and Access) bill*[17], in which the House of Lords pushed to require AI developers to obtain licenses for AI training conducted outside of the UK, as well as to disclose all copyrighted material used to train their models. The House of Lords framed itself as supporting British creators in the process. As a compromise, the government will conduct an economic assessment and publish a report by early 2026[18].

3.2 Implications for AI training

Currently, commercial GenAI training on copyrighted works is not allowed under UK law. While the UK has a strong community of AI researchers and startups, they must either qualify for the non-commercial exception, license all copyrighted works they wish to use for AI training, or train models in other jurisdictions with TDM exceptions.

3.3 Opportunities/Pros

- The UK's narrow TDM exception provides legal certainty for rights holders and AI developers: it makes clear that commercial AI development in the UK requires a license.
- By adopting a TDM exception for TDM research before other jurisdictions, the UK provided additional legal certainty for AI researchers, many of whom call the UK their home. The UK's AI talent has been the basis for AI research labs like DeepMind (later acquired by Google) and the Alan Turing Institute.
- The government's ongoing consultation and commitment to publish an economic assessment by 2026 leaves room for future reform. Policymakers may recalibrate the balance between rights protection and innovation to support commercial AI training while maintaining creator safeguards.

3.4 Challenges/Cons

- The restriction to non-commercial research is highly limiting for industry, as all commercial GenAI training activities are excluded from the exception. This forces developers into time-consuming licensing negotiations or pushes R&D efforts to more permissive jurisdictions, potentially undermining the UK's attractiveness for GenAI investment and talent innovation.
- Regulatory uncertainty persists while the UK government considers possible reforms to introduce broader exceptions or opt-out systems similar to the EU, making long-term planning and resource allocation challenging for both rights holders and technology companies.
- The fragmented licensing landscape and lack of standardized mechanisms for permissions create high transaction and compliance costs, particularly impacting smaller entities lacking negotiating power or technological infrastructure.

[17] Data (Use and Access) Act 2025 <https://bills.parliament.uk/bills/3825#timeline> last accessed on 9th October, 2025

[18] Commentary on AI and copyright: a post-Data Bill UK timeline into 2026 <https://www.pinsentmasons.com/out-law/analysis/ai-and-copyright-post-data-bill-uk-timeline-2026> last accessed on 9th October 2025

- Because the UK is no longer part of the EU market, AI developers may prioritize their investments and compliance in jurisdictions with larger markets and clearer legal regimes for text and data mining, such as the United States or EU.

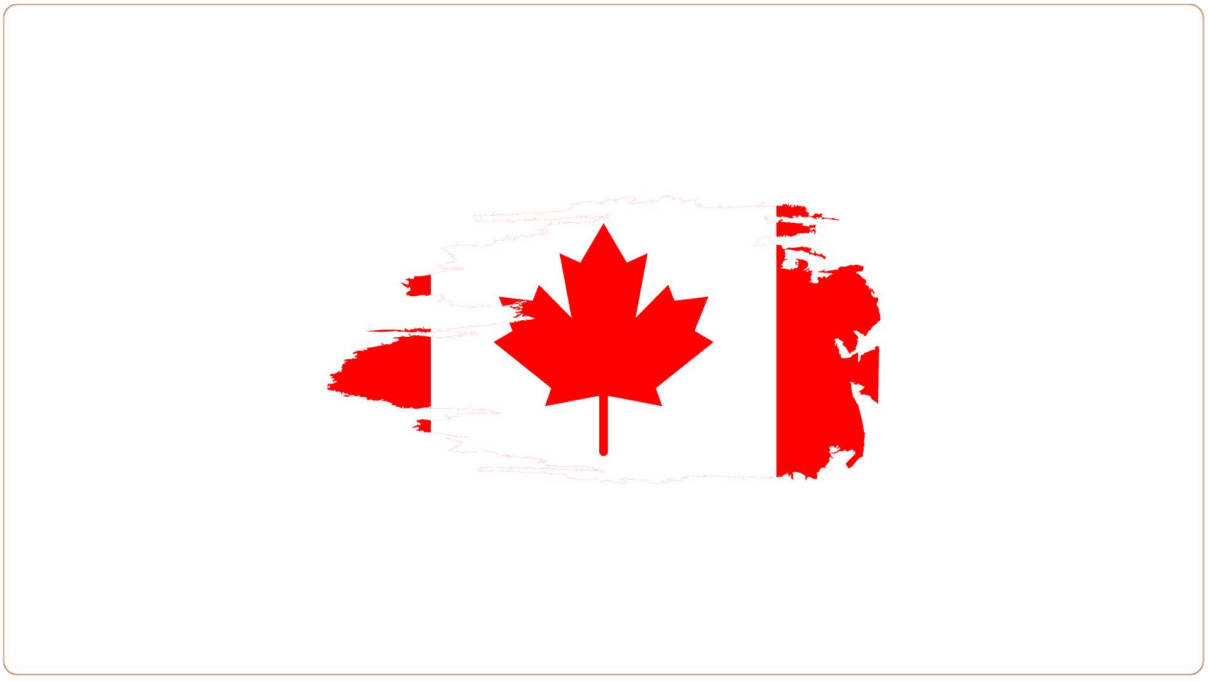
3.5 Key caselaws/developments that are shaping the field

- In the case of Getty Images (US) vs. Stability AI[19], before the High Court of England and Wales, the main claims Getty raised were of infringement under UK copyright law, on the basis of alleged unauthorised use of its photographs in the training and development of Stable Diffusion and in the output that system produced. During the trial, Getty dropped[20] its core copyright claims for training because it faced challenges proving that the infringing acts had taken place in the UK and therefore within the scope of a UK court's jurisdiction. Getty is pursuing separate claims for training in the United States. However, it continues to claim Stability AI is liable for secondary copyright infringement, trade mark infringement, as well as for 'passing off'. The High Court has yet to issue its decision in the case.

3.6 Practical takeaway from UK

The UK's policy is currently characterized by a regulatory deadlock, driven by the highly politicized and emotionally charged conflict between content creators and the AI development sector. While the current regime theoretically mandates negotiated licensing to protect creators and ensure content integrity, the practical reality is that it has unintentionally stifled domestic commercial AI training, leading the activity to be pursued extra-territorially. This means the intended benefit for UK creators ie. the ability to license their works and generate revenue has largely failed to materialize, leaving rights holders feeling unprotected, while simultaneously handicapping the UK's pace of innovation and commercial flexibility. The government's policy process has therefore become stalled, having commenced without a clear strategy or an adequate understanding of the intense stakeholder division, leaving the critical future decision of whether to adopt an EU-style opt-out model or a mandatory licensing framework, as a major unresolved political issue potentially impacting the UK's attractiveness as an AI development hub.

[19] Analysis on Getty Images vs. Stability AI <https://www.pinsentmasons.com/out-law/analysis/getty-images-v-stability-ai-copyright-claims-significance> last accessed on 9th October 2025
[20] *Ibid*



4. Canada: fair-dealing tradition + active policy debate

4.1 What the law is?

In Canada, the Copyright Act[21] lacks a specific TDM exception for AI training, but two general exceptions may potentially apply: the temporary reproduction for a technological process exception and the fair dealing exception. For the former to apply, reproduction must be an essential part of the technological process, exist only for the duration of that process, and solely facilitate a non-infringing use. As a result, if a GenAI system makes temporary copies of copyrighted material that are indispensable for analyzing a large dataset and are temporary, these reproductions could potentially be covered by this exception.

Further, the fair dealing exception[22] permits the use of copyrighted material for specific purposes like research, private study, education, satire, parody, criticism, review, or news reporting, as long as the use is "fair". For criticism, review, or news reporting, source and author attribution is required. This analysis has two parts: first, ensuring that the relevant use qualifies as the specified purpose, and second examining the facts of the use to ensure it is "fair". The fairness of any use is assessed holistically by courts, using factors such as the commercial or non-commercial nature of the purpose, the pattern and scope of the usage, the quantity of copying, available alternatives, the nature of the work, and the effect on the market for the original work. The fair dealing exception is considered a statutory user's right that must be "given a large and liberal interpretation" to balance the public interest and creator

[21] Copyright Act (R.S.C., 1985, c. C-42) <https://laws-lois.justice.gc.ca/eng/acts/c-42/page-6.html> last accessed on 9th October 2025
[22] Copyright Act (R.S.C., 1985, c. C-42) S. 29

rewards, and differs in important respects from U.S. "fair use," which is not applicable in Canada.

In addition to this, between 2023-24, the Government of Canada initiated consultations[23] to modernize copyright laws in the context of AI and the Internet of Things. These consultations revealed sharply divided stakeholder views on how to balance innovation and creator rights. While on the one hand, creators and cultural industries strongly argued that the use of their copyrighted content in TDM without consent and compensation violates their rights; they prefer a voluntary (opt-in) licensing model for TDM and oppose any new copyright exception. They also seek potential downstream remuneration for AI-generated outputs and some support collective management solutions. On the other hand, technology industries and user groups argued that TDM merely involves machine-learning facts and patterns, which should not engage copyright. They supported legal clarifications or a new copyright exception to facilitate TDM, citing the need for large datasets and Canadian competitiveness. They generally oppose licensing as a mandatory first step.

In addition to this, there was widespread support for greater transparency requirements (recordkeeping and disclosure) regarding the copyrighted works used as AI training inputs, largely to allow rights holders to enforce their rights and seek compensation. Technology stakeholders, however, raised concerns that mandatory disclosure could compromise proprietary data, disclose sensitive information, and harm the competitiveness of the Canadian AI industry. The consultation highlighted the need for further fact-finding to better understand the technical process of how copies are made and used during AI training. In September, 2025 the Government of Canada launched an AI Strategy Task Force[24] and initiated public engagement on the development of a renewed AI strategy to position Canada at the forefront of the AI revolution. Public feedback and stakeholder engagement from this will inform Canada's strategy to be an AI leader.

4.2 Implications for AI training?

For AI training and TDM in Canada, the fair dealing exception provides a defense if the usage aligns with an allowed purpose (e.g., research or private study) and meets the fairness factors in practice. There is no blanket legal exemption specifically for AI or TDM; each instance of using protected works for training or analysis must be evaluated for fairness relative to the commercial/non-commercial nature, volume, and impact of the use. Developers of AI systems must be mindful that repetitive or large-scale mining of copyrighted data, especially for outputs that compete with original works or are not attributed where required, may be challenged unless they can convincingly demonstrate that their activities qualify as fair dealing for one of the prescribed purposes. Thus, the legal certainty for AI training in Canada depends

[23] Consultation on Copyright in the Age of Generative Artificial Intelligence <https://ised-isde.canada.ca/site/strategic-policy-sector/en/marketplace-framework-policy/consultation-copyright-age-generative-artificial-intelligence-what-we-heard-report?> last accessed on 9th October 2025

[24] Government of Canada launches AI Strategy Task Force <https://www.canada.ca/en/innovation-science-economic-development/news/2025/09/government-of-canada-launches-ai-strategy-task-force-and-public-engagement-on-the-development-of-the-next-ai-strategy.html> last accessed on 1st December, 2025.

on context and the ability to align TDM practices with the exception's conditions, subject always to judicial interpretation.

4.3 Opportunities/Pros

- The fair dealing exception allows some uses of copyrighted works for research, private study, or education without explicit permission, supporting innovation and academic work, including some AI and TDM activities if deemed fair.
- Ongoing government consultations signal willingness to update law in response to advances in AI, with broad stakeholder engagement suggesting adaptability over time.

4.4 Challenges/Cons

- Lack of a blanket exemption leaves legal uncertainty for AI developers, especially for commercial-scale or repetitive data mining, requiring case-by-case fairness evaluations that may deter innovation or investment.
- There is no clear, easy solution for training data transparency. While greater recordkeeping and disclosure could empower rights holders to enforce their interests, tech companies fear that mandatory transparency will reveal proprietary data or commercially sensitive strategies, reducing Canada's AI sector competitiveness.
- Courts and policymakers have yet to fully address whether machine learning for AI model training qualifies as human "research" under Canadian copyright law, leaving doctrinal ambiguity for AI advances.

4.5 Key caselaws/developments shaping the field

- The Supreme Court of Canada[25] has interpreted "research" broadly ie. it is not confined to non-commercial or private contexts and must be given a liberal construction. Activities aimed at gaining knowledge or understanding, even in commercial settings, can qualify as research under fair dealing. For example: a Supreme Court case[26] recognized that music preview listening (30–90 seconds) to identify user preferences constituted research, showing the doctrine's flexibility. However, no Canadian court has yet ruled on whether AI training using copyrighted material qualifies as "research" within this exception.
- In November 2024, a coalition of Canadian media companies and news publishers initiated an action against OpenAI[27], alleging copyright infringement for using news articles as training data without authorization. Their claim further asserts that OpenAI's practices circumvent technological protection measures and breach the terms of use of various source websites.

[25] CCH Canadian Ltd. v. Law Society of Upper Canada, 2004 SCC 13

[26] Society of Composers, Authors and Music Publishers of Canada v. Bell Canada, 2012 SCC 36.

[27] Alex Riehl, Major Canadian news orgs sue OpenAI for copyright infringement <https://betakit.com/major-canadian-news-orgs-sue-openai-for-copyright-infringement/> last accessed on 9th October 2025

4.6 Practical takeaway from Canada

Canada is in a state of transition. While fair-dealing remains the starting point, legislative reform or industry settlements will likely define commercial GenAI's practical contours.

5. China: regulatory management + copyright compliance requirements

5.1 What the law is?

China's Copyright Law adopts a closed-list approach to exceptions, akin to the EU model, enumerating 13 specific circumstances under Article 24 where works may be used without permission or remuneration, provided attribution is given, normal use is not affected, and rights holders' interests are not unreasonably prejudiced. In the latest revision, Article 24(13)[28] was added as a narrow "catch-all," permitting use in "other circumstances as provided by laws and administrative regulations," which provides flexibility, allowing other statutes or regulations to extend exceptions, leaving open the potential for a specific fair use exception for AI training in future legislative updates.

This provision creates a potential regulatory lever for extending exceptions to activities like AI training through future administrative rules, yet it is not self-executing and requires further legislative or administrative action for specific application. It is important to note here that the

[28] Copyright Law of the PRC (2021 Version) <https://www.chinalawtranslate.com/en/Copyright-Law-of-the-PRC-(2021-Version)/> last accessed on 9th October 2025

practical scope and legal interpretation of Article 24(13) remain undetermined because the provision has not yet been subjected to significant judicial review or litigation.

5.2 Implications for AI training

China's approach to AI and copyright reflects a "moderate leniency" strategy, imposing stricter controls on downstream AI-generated content (AIGC) through measures like mandatory labeling (effective September 2025[29]) and human authorship requirements for protection, while appearing more permissive toward upstream AI training processes. No explicit statutory exception exists for AI training under Article 24 of the Copyright Law, leaving reproduction and adaptation of works during model development legally ambiguous. Yet, China's AI sector remains enormously active, with major players advancing without widespread evidence of licensing copyrighted training data or facing lawsuits specifically targeting the training phase.

This pattern suggests implicit governmental tolerance, prioritizing national AI industry growth, which is deemed strategically vital, over rigorous copyright enforcement at the training stage, akin to protections afforded to other key sectors. Courts have instead focused liability on downstream outputs or contributory infringement, upholding the idea-expression dichotomy without endorsing expansive theories like "retained style."

5.3 Opportunities/Pros

- Strong regulation of downstream AIGC enhances user protection and addresses societal/ethical risks.
- Flexible, post facto regulation adapts with technology without blocking preemptively.
- National standards for security and labeling mitigate harmful or misleading outputs.

5.4 Challenges/Cons

- The focus on post-training content controls and algorithmic filings positions the state as a gatekeeper, which may restrict market access, transparency, and international investment opportunities for local companies.
- Major platforms or companies may more readily comply or influence regulatory processes, potentially consolidating market power and sidelining smaller players and individual creators.

5.5 Key case laws shaping the field

- In another case of **HIPC Ultraman Case (2024, Hangzhou Intermediate People's Court)[30]** involving an AI platform using open-source models, whose users generated images substantially similar to "Ultraman" works, the court found indirect liability (rather than direct), focusing on replacing distribution through AIGC rather

[29] Measures for Identification of Artificial Intelligence-Generated Synthetic Contents <https://www.chinalawtranslate.com/en/ai-labeling/> last accessed on 9th October, 2025

[30] See Hangzhou Intermediate People's Court Civil Judgment (2024) Zhe 01 Minzhong No. 10332. <https://www.kwm.com/cn/en/insights/latest-thinking/chinese-aigc-platform-found-secondarily-liable-for-copyright-infringement.html> last accessed on 9th October 2025

than punishing the underlying technical development. The court held that "Gen-AI technology involves four important stages: data input, data training, content generation and content use. To promote the development of Gen-AI technology, in terms of infringement liability relating to data input and data training stages, a relatively relaxed and tolerant standard should be adopted; in terms of infringement liability relating to AIGC content production and content usage stages, a more stringent liability standard should be adopted…" Later in its opinion, the court also briefly discussed fair use although this case did not actually provoke a fair use argument. The court commented that, "[T]he development of Gen-AI technology requires huge amount of training data in its data input stage, which inevitably requires use of others' [prior] works. The purpose of using others' works in data training stage, in essence, is to learn and analyse the expressions, language characteristics and styles of such prior works, so as to abstract its corresponding rules, structure, models and trends to transform into new works. The purpose of such use…is different from the purpose of representing the originality of the prior work…" This analysis sounds quite similar to the "transformative use" argument adopted by the U.S. courts in their fair use findings under Warhol[31] and Google[32], among other cases. Notably, the court clarified that "in the absence of evidence of intent to use original expressions without authorization or to harm legitimate interests", such training could be considered fair use. This judgement reflects a shift toward more lenient, nuanced judicial assessment, with focus on substitutive harm and the relationship between service providers, trainers, and users.

- While there is still no definitive, binding judicial decision directly on AI training and copyright in China, several landmark cases signal evolving legal reasoning. In the case of **GIC Ultraman Case (2024, Guangzhou Internet Court)**[33]on AI-generated images resembling copyrighted "Ultraman" characters, the provider was found directly liable for infringing the copyright holder's rights of reproduction and adaptation due to substantial similarity in generated images. The court, however, did not require removal of training data, as the defendant had not directly conducted model training. This judgment reflected a strict liability approach, with less distinction between AI training and generative outputs.
- In the case of **BIC REDnote Case (June 2024),**[34] the court examined whether both the process and result of using copyrighted illustrations to train generative models equate to copyright infringement or can be exempted under fair use. The trial details show the growing importance of distinguishing between AI training (upstream use) and AI outputs (downstream content), an issue not yet conclusively addressed by courts.

[31] Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508 (2023)

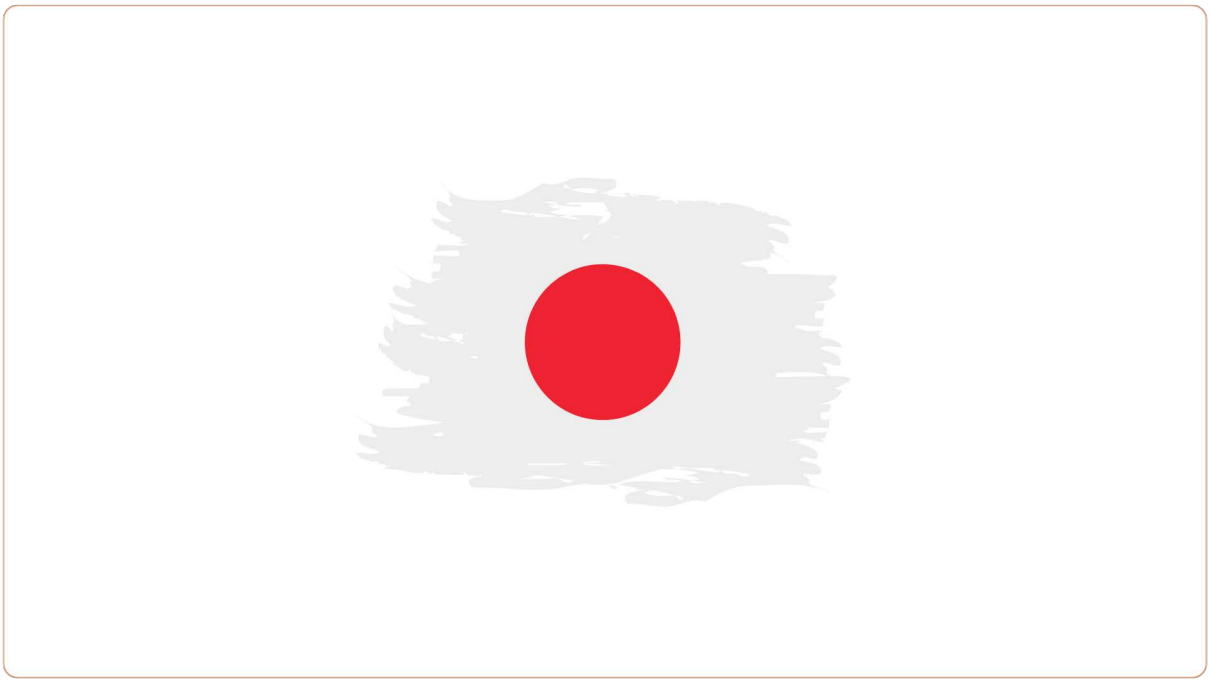
[32] Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015)

[33] See Guangzhou Internet Court Civil Judgment (2024) Yue 0192 Minchu No.113 <https://www.kwm.com/cn/en/insights/latest-thinking/china-s-first-case-on-aigc-output-infringement-ultraman.html> last accessed on 9th October, 2025

[34] See Beijing Intellectual Property Court Civil Judgment (2017) Jing 73 Minzhong No.840.

5.6 Practical takeaway from China

China's approach to AI regulation is anchored in a series of targeted administrative regulations and evolving national standards. Currently, it is operating under a system that mandates theoretical respect for intellectual property but lacks a specific training exception. Regulatory efforts focus strongly on post-training oversight like labeling, security reviews, and content moderation, rather than preemptive restrictions, which helps accelerate industrial innovation. Compliance obligations, such as labeling of AI-generated content and security assessments, are robust yet sometimes favor larger companies better able to navigate regulatory filings. With more focus on downstream accountability, the current approach prioritizes technological advancement over preemptive IP control. In summary, the strategy balances rapid technological growth with rising content accountability and ongoing adaptation to ethical, social, and industrial risks.



6. Japan: comparatively permissive statutory exception for "data analysis"

6.1 What the law is?

Japan's approach to AI and copyright is largely anchored in Article 30(4) of the Copyright Law[35], a 2018 amendment originally designed to facilitate text and data mining (TDM). This provision, which has gained significant attention amidst the rise of GenAI, permits the unlicensed use of copyrighted data for purposes such as testing, data analysis, or data processing. Crucially, the Japanese framework does not explicitly differentiate between legally and non-legally accessed materials, a significant omission compared to the TDM provisions

[35] Copyright Law of Japan <https://www.cric.or.jp/english/clj/cl2.html> last accessed on 9th October, 2025

adopted in the EU, the UK, and Singapore. As a result of this, the law does not expressly prohibit the use of pirated content for AI development.

Further, Japan's Copyright Office issued guidance[36] on AI and copyright earlier this year. The report emphasized that the 30-4 exception for "non-enjoyment use" generally does not harm the economic interests of the author of the copyrighted work used in training. The guidance also noted that "non-enjoyment use" does have limits, such as where GenAI models "overlearn" or simply output copyrighted material.

6.2 Implications for AI training

The Copyright Law in Japan offers a notably broad and permissive legal foundation for AI training. By allowing the unlicensed use of copyrighted data for testing and analysis, the Japanese framework significantly reduces the legal and financial burden on AI developers. The absence of a clear prohibition based on content's source, which is contrasted with the explicit access requirements in the EU, UK, and Singapore, effectively creates an AI development environment with low friction and high legal certainty regarding the inputs. This could be interpreted as a strategic policy choice to accelerate Japan's AI competitiveness, prioritizing technological development over the strict enforcement of copyright holders' exclusive rights concerning the source materials used for data mining.

6.3 Opportunities/Pros

- Operational certainty for text and data miners. The statutory wording has been interpreted to allow broad data analysis, which benefits AI development and allows developers to select data for model training on the basis of what most improves the model's performance or reduces bias. This has also led to a surge in many AI developers opening their offices in Japan, attracting investments into the country.
- Safeguards for creators as it does not cover overlearning or infringing outputs.
- Lower transaction friction than opt-out regimes.

6.4 Challenges/Cons

- Unclear limits on expressive re-use (e.g., whether outputs that reproduce expressive content trigger separate reproduction rights/infringement under the Copyright law).
- Rightsholder discontent: Japanese rightsholders, the news industry in particular, has criticized the law and called for it to be changed.

6.5 Key caselaws/developments shaping the field

- Two major Japanese newspaper publishers, Asahi and Nikkei, have filed a lawsuit[37] against the AI search startup Perplexity, alleging large-scale copyright infringement.

[36] <https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/hoseido/r05_07/pdf/94024201_01.pdf> last accessed on 9th October, 2025

[37] Article on '*Asahi, Nikkei sue Perplexity AI over copyright infringement*' <https://www.asahi.com/ajw/articles/15987899> last accessed on 1st December, 2025

The publishers claim Perplexity's generative AI service illegally copies and stores their copyrighted articles from their servers to generate summaries and answers, effectively engaging in "free riding" and violating their rights of reproduction and public transmission. They assert that Perplexity ignored their explicit robots.txt directives prohibiting unauthorized use. Crucially, the suit also accuses Perplexity of violating the Unfair Competition Prevention Law by generating "answers" that, despite citing the publishers as sources, contain misinformation that deviates from the original articles, thereby damaging the credibility and foundation of journalism. The publishers are seeking an injunction to halt the reproduction of their content and demand its deletion. The matter is currently sub-judice.

- Reflecting on the intellectual property considerations emphasized during the Hiroshima Process, Japan's government has issued formal guidance. The Ministry of Economy, Trade and Industry (METI), in its AI Guidelines For Businesses[38] published in 2024, explicitly advised Japanese firms to be diligent. The guidelines recommend that businesses "Implement appropriate data input measures…to honor rights related to privacy and intellectual property, including copyright-protected content." This underscores the commitment to responsible IP practices within the country's AI sector.
- In response to the Japanese government's "IP in the AI Era" consultation, the Nippon Anime and Film Culture Association (NAFCA) put forth several key recommendations[39]. The association specifically urged the government to implement watermarking as a mechanism to identify material generated by AI. Furthermore, NAFCA called for stronger protection of neighboring rights, such as the right of publicity, to prevent the exploitation of recognizable voice actors by GenAI technologies.

6.6 Practical takeaway from Japan

Japan's regulatory framework for AI is distinguished by a clear policy decision that fundamentally separates data analysis, which encompasses AI training, from traditional copyright-restricted uses. The government has explicitly recognized that the ingestion of copyrighted works for the purpose of data analysis does not constitute a use that "enjoys" the artistic or expressive elements of the work. This pragmatic stance effectively distinguishes AI training as a distinct category of use that does not directly harm the core market or interests of rights holders typically protected by licensing, thereby encouraging technological development.

This interpretation is codified in a statutory exception that allows firms significant confidence to use copyrighted works for large-scale AI training within Japan. However, this progressive stance on upstream training does not resolve the many subsequent legal and commercial disputes arising from AI outputs. This ambiguity is evident in the ongoing conflicts with the news media sector, where publishers accuse AI search products of undermining their business model by using internet content to generate answers, demonstrating that while the training

[38] AI Guidelines for Business <https://www.diligent.com/resources/blog/japan-ai-regulations> last accessed on 9th October, 2025

[39] Public Comments <https://nafca.jp/public-comment01/> last accessed on 9th October, 2025

input is legally permissive, the downstream impact on content consumption and attribution remains a complex, unresolved challenge.

7. Singapore: Forward-looking approach while balancing interests

7.1 What the law is?

Under the Singapore Copyright Act 2021,[40] there is a statutory exception allowing the copying of copyright works for the purpose of "computational data analysis" (CDA). This exception is found in Section 244[41] of the Copyright Act 2021, which defines "computational data analysis" as "the use of a computer program to identify, extract, and analyze information or data from a work or recording, including using the work as an example of a type of information or data to improve the functioning of a computer program related to that type of data." The Intellectual Property Office of Singapore (IPOS) has clarified that this includes (a) Text and data mining (TDM), (b) Sentiment analysis and (c)Training machine learning models, including GenAI.

To rely on this exception, users must meet several conditions including (a) lawful access, ie., the work must be lawfully accessed (e.g., via a licensed database or legally obtained copy), and (b) non-sharing rule ie. copies made for CDA cannot be shared with others, except for verifying results, or for collaborative research or study relating to the CDA purpose. Further, the law prescribes that copies must not be used for any other purpose beyond CDA and the

[40] Copyright Act 2021 <https://sso.agc.gov.sg/Act/CA2021?ProvIds=P15-#pr244-> last accessed on 9th October, 2025

[41] Copyright Act, 2021 s. 244

work copied must not be an infringing copy unless it's necessary for a prescribed analysis, or the user did not and could not reasonably know it was infringing. It is also important to note here that as per S. 187[42] of the Act, the law makes void any contractual terms that prohibit using the exception under 244 ie., copyright owners cannot prohibit data analysis in a contract to make access "unlawful".

7.2 Implications for AI training

The CDA exception allows AI developers and researchers in Singapore to train GenAI systems without infringing copyright, provided they comply with the conditions on lawful access and restricted use of copies. This legal clarity encourages investments in AI innovation and supports Singapore's national AI strategy, positioning the country as a leader in AI development. It means that training datasets can include copyrighted materials accessed lawfully without explicit permission from copyright holders. Nonetheless, there remains a risk that outputs generated by AI based on such training could infringe copyright if they closely replicate copyrighted works. Thus, while input training is protected, output must be carefully managed to avoid infringement.

7.3 Opportunities/Pros:

- Enables innovation by allowing extensive use of copyrighted data for AI training without infringement liability.
- Covers commercial use, attracting investment and supporting Singapore's AI ecosystem.
- Provides legal certainty about permissible AI training methods.
- Supports a wide range of data analysis techniques like TDM, sentiment analysis, and machine learning improvements.

7.4 Challenges/Cons:

- AI-generated outputs still pose copyright risk, requiring further legal or policy clarity.
- Possible ambiguity and enforcement challenges regarding what constitutes lawful access and infringing copies.

7.5 Key caselaws/developments shaping the field

Singapore has not yet seen major litigation specifically testing this exception in AI contexts. However, the legal framework is proactive and among the first in Southeast Asia to address AI-related copyright issues comprehensively. Model Governance Framework[43] published by Singapore on GenAI is insightful on the government's views of the relationship between copyright and AI. The framework proposes that access to data needs to be "balanced" with interests in copyrighted works under the pillar of fostering a trusted AI ecosystem. Given the

[42] Copyright Act, 2021 s. 187

[43] Model AI Governance Framework for Generative AI <https://aiverifyfoundation.sg/wp-content/uploads/2024/05/Model-AI-Governance-Framework-for-Generative-AI-May-2024-1-1.pdf> last accessed on 9th October, 2025

large amounts of data needed to train state-of-the-art GenAI models, the guide suggests developing "approaches" to resolve conflicts in a "clear and efficient manner." Internationally, Singapore's approach is more liberal compared to jurisdictions like the EU where commercial use exceptions are more limited, positioning Singapore as an attractive hub for AI development.

7.6 Practical Takeaway from Singapore

Singapore's computational data analysis exception in copyright law creates a forward-looking balance between protecting copyright owners and enabling AI innovation through lawful access to copyrighted works for machine learning. This legal carve-out fosters a vibrant AI ecosystem by allowing training on copyrighted data under controlled conditions but underscores the continuing need to manage copyright risk in AI-generated outputs. This policy approach attracts several AI developers to set-up offices/employee bases in the country. Singapore's approach serves as a model for jurisdictions aiming to support AI while respecting intellectual property rights in an evolving digital economy.

III. Policy Prescriptions

This concluding section translates comparative lessons from various global jurisdictions into practicable, India-focused policy prescriptions. It targets three core stakeholder groups, AI policymakers, AI developers, and AI rights-holders & cultural industries, and the aim is to balance innovation and investment in AI with a predictable, enforceable rights framework that protects creators and public interest objectives (access, competition, privacy and cultural diversity).

Sr No.	Policy Prescription	Rationale
Prescriptions for AI Policymakers		
1.	**Enact a Broad, Technologically Neutral TDM Exception:** India should legislate a broad statutory TDM exception covering both commercial and non-commercial AI training. This reduces the risk of legal challenges under the current, narrow "fair dealing" provision[44], which is ill-suited for modern, large-scale data processing.	This provides the legal certainty and low barrier to entry necessary for startups and large enterprises to access vast public data, mirroring the permissive environment of Japan/Singapore, and making Indian AI development globally competitive by enabling scale and speed without the legal ambiguity currently present.
2.	**Voluntary Opt-Out Registries for Creators:** Policymakers should explore the feasibility of establishing a nationally recognized, voluntary, machine-readable registry where creators (both individual and corporate) can register their works and explicitly reserve their TDM rights.	This provides individual creators, particularly those lacking collective bargaining power, with a low-cost mechanism to assert control without requiring developers to seek millions of licenses. By being voluntary, it avoids the chaos of a mandatory, comprehensive opt-out system while giving developers a clear list of exclusion zones to legally mitigate risk.
3.	**Voluntary Attribution Obligations in Generated Outputs:** The government can explore how commercial Generative AI systems, where technically feasible, include an attribution watermark or metadata tag when outputs are substantially similar to works in the training data, or when responding to a specific user prompt that clearly references a source.	This fosters transparency and ethical development. It helps consumers understand the provenance of the AI output and provides creators with the necessary attribution for the use of their works.
4.	**Promote Metadata Standards:** Given the vast volume of copyrighted	This is a light-touch regulatory approach that minimizes compliance overhead for

[44] Indian Copyright Law (1957) <https://copyright.gov.in/documents/copyrightrules1957.pdf> last accessed on 9th October, 2025

	works used in AI training datasets, manual clearance or opt-in licensing is often impractical. Machine-readable metadata, such as standardized rights statements, automated attribution tags, and reserved-use markers, could allow automated systems and web crawlers to efficiently identify works subject to specific copyright restrictions or opt-outs. This facilitates responsible data sourcing and helps prevent unintentional infringement by respecting content creators' rights without stifling innovation.	developers while providing the building blocks for balancing creators' rights, thereby maintaining a fast pace of development.
5.	**Adopt a Guidance-First Approach to IP:** Use Ministry guidelines and advisories to encourage "best effort" compliance regarding pirated content, reserving hard-coded law for high-risk IP violations (e.g., when there is a direct case made out for market harm/displacement).	This reflects the government's stated pro-innovation stance, ensuring that the regulation is proportionate to the risk and does not create an immediate chill on research and experimentation by penalizing unintentional or secondary infringement.
6.	**Output Filters (Voice and Image Protection):** Government can also explore specific guardrails that require technical filters or restrictions on the use of protected identifiers (such as a recognizable voice actor's voice or a living person's likeness) to prevent unauthorized commercial exploitation (deepfakes or voice cloning) without explicit consent. This can be explored by enacting tailored digital replica or anti-deepfake laws focusing on preventing impersonation, fraud, harassment (e.g., false explicit images), and privacy harms unless explicit consent is given or free expression interests justify the depiction. Such laws should resemble privacy frameworks by outright banning non-consensual conduct, while for commercially valuable digital replicas of famous individuals, carefully defining licensing	This helps to protect "neighboring rights" and the right of publicity, addressing one of the most acute and ethically charged harms faced by performers and recognizable public figures in India without restricting the TDM phase itself.

	mechanisms, consent standards, and remuneration to balance rights holder protections with innovation. Technical filters can enforce these rules by restricting generation of such content absent verified permissions.	
Prescriptions for AI Developers		
7.	**Implement Robust Dataset-Provenance and Self-Certification:** Developers must maintain an internal, auditable log of dataset provenance. This should be used for internal risk management and voluntary disclosure to build trust with investors and global partners.	This de-risks Indian AI models for global deployment. By having clear proof of legal sourcing, developers in India can effectively counter legal challenges in markets like the US and EU, ensuring that domestic innovation is export-ready and resilient.
8.	**Adopt Defensive Data Filtering Against Piracy:** Developers should proactively filter out content from clearly identified pirated sources.	This maintains an ethical and legal clean slate. Avoiding pirated inputs minimizes the risk of courts viewing the development process unfavorably which happens to be a key factor in copyright litigation, thereby protecting the long-term viability and reputation of the company.
9.	**Prioritize Strategic, Selective Licensing:** Engage in strategic voluntary licensing with key Indian content owners (e.g., major news publishers, regional film studios) where the risk of legal challenge is highest.	This stabilizes the domestic market by turning potential adversaries (large rights holders) into partners, reducing friction and securing access to *exclusive* Indian cultural datasets that provide a unique competitive advantage to local models.
Prescriptions for AI Right Holders' and Cultural Industries		
10.	**Focus on Output Harm, Not Input Prohibition:** Creators should shift their primary legal focus to the output of AI systems. Litigation and policy efforts should prioritize cases where the output is a clear substitution for, or reproduction of, the original human work, demonstrating direct market harm to challenge the scope of *fair dealing*.	This is a more effective legal strategy in most jurisdictions. By focusing on market displacement, rights holders can secure damages and injunctions that protect their core business model without trying to block the foundational and legitimate TDM process.
11.	**Implement Machine-Readable Rights Metadata:** Content owners	This provides creators with granular control over their content in the digital

	should be incentivized to embed machine-readable opt-out reservations within the metadata of their published works.	environment, enabling both effective enforcement of their opt-outs and the seamless, automated negotiation required for the scale of the AI economy.
12.	**Mutual Cooperation for Novel Product Development:** Rightsholders and AI companies should prioritize identifying areas of mutual cooperation to create innovative products that were technologically infeasible before the advent of AI. This involves shifting the focus from litigation to co-creation. This strategy allows rightsholders to leverage their unique creative assets in new ways, while AI developers gain access to high-quality, legally-cleared training data and novel product verticals. For eg., the news industry could experiment with AI search best practices, such as designing approaches to attribution/citations that encourage click-through or ensure users who are getting a quick, factual answer see a publisher's brand and build more mindshare.	This will help center the economic model from conflict to collaboration. AI developers urgently need high-quality, legally clear data, which rightsholders possess in the form of unique archives and creative works. This collaboration allows cultural industries to leverage previously undervalued assets by transforming them into new, interactive, and personalized AI-driven experiences, ensuring their long-term relevance and financial stability.

Author

JAMEELA SAHIBA

Associate Director- AI and Public Affairs, The Dialogue

Jameela Sahiba manages the Secretariat of the Coalition for the Responsible Evolution of AI (CoRE-AI) and leads The Dialogue's work on Artificial Intelligence and public affairs. She coordinates engagement with Parliamentarians and government stakeholders on technology policy. She holds a law degree from Symbiosis Law School, Pune. Prior to her current role, she served as Chief of Staff to Dr. Amar Patnaik, Member of Parliament (Rajya Sabha), where she worked on legislative matters and stakeholder outreach. She has over eight years of experience in public policy, with a focus on emerging technologies and legislative processes.

thedialogue.co

@_DialogueIndia

@TheDialogue_Official

@The-Dialogue-India

@TheDialogue