

BUILDING SANDCASTLES AT HIGH TIDE: CONTRACTS IN THE IP REGULATION OF AI TRAINING DATA

Contracts have been proposed to complement copyright law to better regulate the use of data in training AI models. Still, the dynamic between contracts and copyright law in this context remains underexplored. This article examines their complementary and conflicting dynamics, identifying an “emerging contract override” across various scenarios, which includes how contracts interact specifically with copyright exceptions, how contracts serve as a mechanism to create rights, their role in the opt-out and opt-in debate, and their interaction with the transparency requirement. The analysis shows that regulating AI data through contracts is fragmented, temporary, and uncertain – like building a sandcastle at high tide. While contracts enable tailored private ordering, development in copyright law, competition law, and technological advances can easily erode their enforceability and coherence.

Wenting **CHENG**

*PhD (The Australian National University), MPhil (Peking University);
Senior Lecturer, School of Law, the University of Queensland.*

Georg **ZOELLER**

Chief Strategist/Co-Founder, Centre for AI Leadership (C4AIL).

I. Introduction

1 The use of data for training artificial intelligence (“AI”) models raises significant concerns about potential copyright violations.¹ Contracts, in particular, collective licensing agreements, have been recognised and proposed as one of the most effective alternative tools to balance the rights of copyright holders and creators on the one hand, while addressing legal challenges and accelerating the development of AI technologies on the other.² The role of contracts is evident in

1 Benjamin Sobel, “Elements of Style: Copyright, Similarity, and Generative AI” (2024) 38 *Harvard Journal of Law & Technology* 49.

2 This has been proposed by academic work. see Nicola Lucchi, “ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems” (2023) 15 *European Journal of Risk Regulation* 602. See also UK Intellectual Property (*cont’d on the next page*)

their complement to copyright, as they regulate the downstream transferability of copyrighted material through licensing agreements that specify certain requirements for the use of copyrighted material. Furthermore, as AI-related technology continues to evolve, it is crucial for the law to keep pace with its advancements. Copyright law, as a statutory law, may be slow to be amended, and contracts can incorporate flexible rules that are capable of adapting to new technological realities.

2 Contracts can be an important instrument of private regulation and have played a fundamental role in the copyright equation. Historically, copyright assignments between booksellers kept the market brimming with works of all types and contributed to the continuing vitality of works out of copyright. The network of contracts between British booksellers and related downstream aspects of author-bookseller contracts formed the legal foundation for the widespread distribution of works of authorship.³ In terms of subject matter, there are overlaps, as both are regulatory instruments for specific information that can be fixed in a tangible medium of expression. In most cases, copyright law defines the rights and contracts (licensing agreements) that facilitate the enjoyment of the right through authorisation. On the other hand, contracts also complement copyright law when the law is not clear about specifics in the content under protection or the rights conferred.

3 Since contracts may govern aspects that copyright does not regulate and can sometimes override what copyright law provides, there may be a tension between the two. Prominently, the objective of contracts as a regulatory instrument may conflict with other laws, as pointed out by a 2010 UK Intellectual Property Office report (“2010 UK IPO Report”) on the relationship between copyright and contract law:⁴

There are obvious conflicts between the aims of competition law, the aims of copyright law to provide an incentive by excluding competition, the aims of *droit d’auteur* inspired legislation to secure a livelihood for creators and protect their personality, and general principles of efficient contracting that may or may not conflict with other notions of fairness. The regulation of contracts may also

Office, *Copyright and Artificial Intelligence: Consultation* (CP 1205, 17 December 2024) <https://assets.publishing.service.gov.uk/media/6762c95e3229e84d9bbde7a3/241212_AI_and_Copyright_Consultation_print.pdf> (accessed 11 September 2025). Also see US Copyright Office, *Copyright and Artificial Intelligence – Part 3: Generative AI Training* (6 May 2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>> (accessed 23 September 2025).

3 Robert W Gomulkiewicz, “Contracts Mattered as Much as Copyrights” (2018) 66 *Journal of the Copyright Society* 441.

4 Martin Kretschmer *et al*, *The Relationship Between Copyright and Contract Law* (1 July 2010) Intellectual Property Office (UK) Research Paper No 2010 (04) at p 9.

have to take note of fundamental rights and freedoms (such as constitutional principles and human rights).

4 Despite the general understanding of contracts as a tool of private regulation in relation to copyright, in the specific area of AI training data, the role of contracts as well as their feasibility or effectiveness, is not sufficiently understood in the literature. This is partly because contracts, as voluntarily assumed legal obligations, are only binding on the parties. Every contract is contingent on the context, making it challenging to generalise how contracts as a regulatory tool have worked. It is not up to the state to dictate how contracts should work either – after all, if the parties themselves make a bad or vague bargain, the court would be reluctant to fix it, particularly if the parties themselves knowingly entered into it. The effects of contract regulation are also difficult for outsiders to discern. Most contracts are quietly performed and discharged, and only those brought to court would draw public attention.

5 This is true for contracts as a regulatory tool for data used in AI training. Most platforms nowadays have updated their terms of service to include a clause about using their content for AI training. There are very few studies dedicated to the role of contracts (more specifically, contractual terms) in regulating data used for AI training. Among existing literature, Henderson and Lemley pointed out that “no model creator has actually tried” to enforce the contractual terms because the authors think legal enforceability of these licenses is questionable.⁵ The first case in the US using breach of contract as a cause of action occurred soon after the paper was published. In June 2025, Reddit sued Anthropic for breaching its terms of service by commercially exploiting the platform. The relevant agreement reads: “You may not, without our written agreement: license, sell, transfer, assign, distribute, host, or otherwise commercially exploit the Services or Content.”⁶ The cause of action also involves unjust enrichment, trespass to chattels, and unfair competition. This may indicate a new trend in litigation that is diverting from focusing on copyright infringement. In addition, the recent US Copyright Office’s report on generative AI training finds that voluntary licensing of copyrighted works for use in AI training is increasingly occurring in various sectors between AI developers and copyright owners.⁷

5 Peter Henderson & Mark A Lemley, “The Mirage of Artificial Intelligence Terms of Use Restrictions” (2025) 100 *Indiana Law Journal* 2.

6 *Reddit Inc v Anthropic PBC* CGC-25-625892 (4 June 2025, Super Ct Cal) (US).

7 US Copyright Office, *Copyright and Artificial Intelligence– Part 3: Generative AI Training* (6 May 2025) at p 85 <<https://www.copyright.gov/ai/Copyright-and-Training>> (cont’d on the next page)

6 Despite the progress in reality, there are still many thorny issues to address, such as whether datasets are copyrightable, who owns the data, whether scraping behaviour can be governed by contract, and how. Most importantly, it is not clear how the dynamics between copyright and contracts work and what factors may impact using breach of contract as a cause of action in relation to AI training data.

7 This article examines how contracts function as a regulatory tool for data used for AI training purposes. Specifically, it addresses a crucial question: What role do contracts play in regulating the data used for AI training? The relationship between copyright and contracts is discussed at a more nuanced level, illustrating how their general dynamics are present in the context of AI training data. In addition to outlining that they are complementary and sometimes conflicting with each other, this article categorises situations of copyright-contract conflicts, which are enunciated as specific scenarios of the emerging contract override. These scenarios include: (a) the interaction of contracts with copyright exceptions (particularly fair use); (b) the use of contracts as a rights-creation mechanism; (c) the role of contracts in regulating creator rights; and (d) the interaction of contracts with the transparency requirement.

8 Given the complexity of identified issues, a comprehensive scan of the issues in major jurisdictions is beyond the scope of this article. The selection of illustrations is based on the latest judicial and legislative developments available from the literature that may serve this article's categorial analysis. For instance, US and China are used as examples to discuss the copyright-contract conflict. In the US, federal courts have addressed this tension through the Copyright Act's express pre-emption provision, which prohibits states from creating rights that are equivalent to copyright.⁸ In China, contractual provisions that contradict mandatory provisions in laws, regulations, and administrative rules are invalid.⁹ Despite the jurisdictional difference, these examples can be

Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf> (accessed 23 September 2025).

- 8 Guy A Rub, "Moving From Express Preemption to Conflict Preemption in Scrutinizing Contracts Over Copyrighted Goods" (2023) 56 *Akron Law Review* 303.
- 9 Civil Code of the People's Republic of China (effective 1 January 2021) Art 153 para 1 provides: "A civil juristic act that violates the mandatory provisions of laws or administrative regulations shall be invalid. However, this shall not apply if the mandatory provisions do not render the civil juristic act invalid." See Wang Liming (王利明), "On the Determination of the Invalidity of Contract on the Basis of Normative Purpose: with Article 16 of the Interpretation on General Provisions of Contracts in the Civil Code as the Center" (论依规范目的认定合同无效: 以《合同编通则解释》第16条为中心) (2024) *China Legal Science* (中国法学) 24. Wang
(cont'd on the next page)

considered illustrations of a broader contract law doctrine of illegality – that contracts cannot lawfully require parties to act in contravention of statutory duties.¹⁰ This article further illustrates how the principle of illegality has general implications for copyright-contract interaction by referring to cases in other common law jurisdictions.

9 It is argued that regulating AI training data *via* contract is like building a sandcastle at high tide. The regulation is inherently individualised and fragmented. While contracts can take into account specific considerations, they cannot create a fortress to achieve concerted regulatory objectives. While the copyright-contract conflicts entail considerable uncertainty, the doctrine of illegality suggests the temporality of contractual regulation – the clarification of the issue by copyright law can render inconsistent contract terms unenforceable. In addition, other headwinds are blowing down the contract sandcastle, including considerations of advanced technology, competition law, and consumer law. While contracts manifest private ordering at work for AI training data, the sandcastle metaphor reveals its function is only temporary beauty, rather than durability.

10 This article contributes to the literature by filling a couple of gaps regarding contractual regulation of AI training data. Firstly, there is a lack of discussion on the *status quo* to understand how contracts have actually played a role in regulating AI training data, in particular, from the dynamics between copyright and contracts. Copyright-contract interactions become nuanced and complicated when there is a conflict between the objectives of copyright law and contracts (as identified in the 2010 UK IPO Report above¹¹), and when such a conflict is often addressed through technicalities to determine which set of rules prevails in a specific context. The authors have undertaken a comprehensive survey of the existing policies, guidelines, and case law in major jurisdictions to provide an up-to-date overview. For instance, it is useful to note that *Reddit Inc v Anthropic PBC*¹² was initiated during the writing process.

argues that the phrase “if the mandatory provisions do not render the civil juristic act invalid” in the second half of the above provision should consider the nature of mandatory provisions, combined with the purpose of the regulation.

10 See, eg, the UK case *Patel v Mirza* [2016] UKSC 42. The case shifted from the traditional rule-based approach to the doctrine of illegality, to a discretionary policy-based analysis that considered whether the enforcement of the contract might undermine the purpose of the relevant statute. See also *Ochroid Trading Ltd v Chua Siok Lui* [2018] 1 SLR 363. For details, see Part III.C of this article.

11 Martin Kretschmer *et al*, *The Relationship Between Copyright and Contract Law* (1 July 2010) Intellectual Property Office (UK) Research Paper No 2010 (04).

12 CGC-25-625892 (4 June 2025, Super Ct Cal) (US).

This provides an insight into the *status quo*, while recognising that it is a fast-evolving area.

11 Secondly, while there is a discussion about the relationship between copyright and contracts in different jurisdictions (in the US this is known as the copyright pre-emption doctrine),¹³ not much has been said about how this issue has been addressed in relation to the contractual regulation of AI. The existing literature has briefly discussed the role of contracts in regulating input and output data, without drawing on the discussion of copyright-contract dynamics. For instance, it is observed that regarding the issue of intellectual property protection for input data, contracts, especially licensing agreements, can be helpful tools for obtaining permission to use data for AI training purposes and for addressing related concerns. However, when it comes to IP protection of AI-generated outputs, contracts often serve as a means for AI operators and developers to monopolise these outputs, often without proper justification.¹⁴ While conclusions like this shed light on the discussion in general, it is not clear how the traditional relationship between copyright and contracts is applied in relation to the regulation of AI training data. This article contributes an analysis of the copyright-contract dynamics both broadly and specifically.

12 Before proceeding further, the scope of the article needs to be defined. The focus is on the role of contracts in regulating AI training data, in relation to intellectual property. Regarding the issue of intellectual property protection for AI-related content, a distinction is often made between input data and output data. Input data refers to the information used to train AI models, which often includes text, video, audio, and their combinations. Text data, such as news and books, social media posts, and code repositories, taking into consideration language style and topic diversity, can be used for tasks like language modelling, translation, summarisation, question answering, sentiment analysis, and code generation. Visual data, including photos, satellite images, medical scans, and CCTV footage, are often used in models that aim to help with object detection, segmentation, facial recognition, action recognition, and medical diagnosis. Audio data (including speech, music, and environmental sounds) can be used for models tasked with automatic

13 Mark A Lemley, "Beyond Preemption: The Law and Policy of Intellectual Property Licensing" (1999) 87 *California Law Review* 111; Guy A Rub, "Copyright Survives: Rethinking the Copyright-Contract Conflict" (2017) 103 *Virginia Law Review* 1141; Guy A Rub, "Moving From Express Preemption to Conflict Preemption in Scrutinizing Contracts over Copyrighted Goods" (2023) 56 *Akron Law Review* 303.

14 Nataliia Filatova-Bilous, "Intellectual Property, Artificial Intelligence and Contract Law: Are There Any Intersections and Fruitful Collaborations Between These Areas?" (2024) 6 *Visegrad Journal on Human Rights* 33.

speech recognition, speaker identification, music transcription, sound event detection, and text-to-speech applications. There are also sensor data, structural data, and time series signals (such as weather data). The copyright status of these data varies from case to case, and most of them may not be copyrightable. On the other hand, output data are those generated by AI models. While the copyright status is also uncertain at the time of writing, the US Copyright Office report has clearly provided guidance for the copyrightability of AI-generated data.¹⁵ Furthermore, contracts are widely used to regulate other aspects of AI, including privacy issues, personal data protection, and competition-related issues.¹⁶ These are also beyond the scope of this article.

13 There are also two related issues that will only be mentioned tangentially when relevant to the discussion. Firstly, the copyrightability of the subject matter (AI training data) is a crucial issue for this article. The authors' discussion is based on general practice at the time of writing. However, copyright issues will not be discussed in much detail as there has been emerging literature on this topic. Secondly, there is much to be discussed regarding the role that contracts play in regulating output data generated by AI.¹⁷ For present purposes, this article will mainly focus on the role of contracts in regulating input data.

II. Contracts complement copyright?

14 When AI training data are clearly copyrightable, the relationship between contracts and copyright is not so different from their relationship in the traditional sense – copyright defines the title, and contracts regulate transferability of the title. This complementary relationship is built on the basis that copyright is a statutory right that defines the rights conferred to copyright holders, along with its limitations and exceptions. Contracts, often in the form of licensing agreements, can further provide how the rights holder permits others to use their rights upon agreement of parties. These include further clarification of rights and obligations of parties, the purpose, scope and terms of the authorisation, the manner of remuneration (royalties), and remedies

15 US Copyright Office, *Copyright and Artificial Intelligence– Part 3: Generative AI Training* (6 May 2025) <<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>> (accessed 23 September 2025).

16 Peter Henderson & Mark A Lemley, “The Mirage of Artificial Intelligence Terms of Use Restrictions” (2025) 100 *Indiana Law Journal* 2.

17 Readers may wish to consult other contributions in this Special Issue on Artificial Intelligence and Intellectual Property in the Asia-Pacific for complementary perspectives on this topic.

for breach of contract, and other procedural rights, such as arbitration clauses and jurisdiction of disputes. When AI training data (such as musical works) are copyrightable, licensing agreements for the use of such works for the purpose of training may differ from traditional contracts in specific terms of permission. Nonetheless, the role that contracts play here is the same as that of traditional licensing agreements, and thus is not the focus of this article.

15 However, controversial issues relating to AI training data may challenge the complementary relationship between contracts and copyright. Firstly, in some circumstances, training data *per se* may not be copyrightable, often due to the concern that these data would not meet the originality requirement of copyright law.¹⁸ Model weights and outputs may also not be copyrightable.¹⁹ This means that the traditional licensing agreements may not work well here. Nonetheless, the copyright laws on this issue vary from jurisdiction to jurisdiction.²⁰ If, based on *de facto* control (or nothing), the data holder incorporated a clause that prohibits the use of data by a third party without specific permission in the contract, the question then becomes: To what extent does such a clause have legal effect? If breach of such a term itself can become a cause of action, does that mean that contracts can, in these circumstances, create rights that are not yet available in statutes? This issue will be analysed in detail in Part IV.B.

16 Secondly, there have been arguments that even though the material used for training AI is copyrightable, such use does not necessarily constitute infringement. For instance, using copyrighted works for AI training can be defended under the fair use/fair dealing doctrine under copyright law, which may permit limited use of copyrighted material without requiring consent from rights holders. This has been argued based on law and economics – under the fair use doctrine, no infringement may be established if the model’s outputs extend beyond the original purpose of the inputs.²¹

17 There are also doctrinal arguments in support of applying the fair use doctrine based on copyright law.²² However, it should be noted that the doctrine has evolved differently in different jurisdictions, and

18 Pamela Samuelson, “Generative AI Meets Copyright” (2023) 381 *Science* 158.

19 Peter Henderson & Mark A Lemley, “The Mirage of Artificial Intelligence Terms of Use Restrictions” (2025) 100 *Indiana Law Journal* 2.

20 Adam Buick, “Copyright and AI Training Data – Transparency to the Rescue?” (2024) 20 *Journal of Intellectual Property Law & Practice* 182.

21 Mark A Lemley & Bryan Casey, “Fair Learning” (2020) 99 *Texas Law Review* 743.

22 Jessica L Gilotte, “Copyright Infringement in AI-Generated Artworks” (2019) 53 *UC Davis Law Review* 2655.

even different terminologies are adopted – fair use in the US and fair dealing in other common law jurisdictions.²³ The difference is that fair dealing involves a two-stage analysis. First, the court needs to decide whether the intended use qualifies for one of the specified purposes, and then the court will analyse whether the use is fair. But the US fair use doctrine only considers the second-stage question.²⁴ Since American fair use can potentially apply to any purpose²⁵ without explicit statutory limitations, there have been significant jurisprudential developments relying on the doctrine of fair use, primarily found in recent US cases, including *Bartz v Anthropic PBC*²⁶ (“Bartz”), *Kadrey v Meta Platforms Inc*²⁷ (“Kadrey”) and *Thomson Reuters Enterprise Centre GmbH v ROSS Intelligence Inc*²⁸ (“Thomson Reuters”). In *Bartz*, Anthropic used both lawfully purchased books and pirated books to train its large language model, Claude. Fair use was upheld for lawfully purchased books due to their highly transformative nature. But for pirated books, the fair use argument was rejected due to their inherently and irredeemably infringing nature. In *Kadrey*, fair use was supported due to the plaintiff’s failure to provide evidence of market harm. But the court warned that future plaintiffs in similar cases with stronger evidence of stronger market harm may well prevail. In *Thomson Reuters*, where a non-generative AI legal search tool used Westlaw headnotes without a license, the fair use defence was rejected mainly because the use was non-transformative and directly competitive, which harmed the plaintiff’s existing and potential markets.²⁹

18 Fair dealing in other common law jurisdictions (UK, Canada, and Australia) is narrower. For instance, in Australia, the use would

23 Ariel Katz, “Debunking the Fair Use vs. Fair Dealing Myth: Have We Had Fair Use All Along?” in *The Cambridge Handbook of Copyright Limitations and Exceptions* (Shyamkrishna Balganes, Ng-Loy Wee Loon & Haochen Sun eds) (Cambridge University Press, 2021).

24 Michael Geist, “Fairness Found: How Canada Quietly Shifted From Fair Dealing to Fair Use” in *The Copyright Pentology: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law* (Michael Geist ed) (University of Ottawa Press, 2013).

25 This is provided in Copyright Act 17 USC (US) § 107 (2019), including: (a) the purpose and character of the use; (b) the nature of the copyrighted work; (c) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (d) the effect of the use upon the potential market for or value of the protected work.

26 No 3:24-cv-05417 (23 June 2025, ND Cal) (US).

27 No 3:23-cv-03417 (25 June 2025, ND Cal) (US).

28 No 1:20-cv-613-SB, 2025 WL 458520 (11 February 2025, D Del) (US).

29 For comments on implications, see David M McIntosh *et al*, “A Tale of Three Cases: How Fair Use Is Playing Out in AI Copyright Lawsuits”, *Ropes & Gray* (15 July 2025) <<https://www.ropesgray.com/en/insights/alerts/2025/07/a-tale-of-three-cases-how-fair-use-is-playing-out-in-ai-copyright-lawsuits>> (accessed 23 September 2025).

have to first fall into the explicit purposes enumerated in the statute,³⁰ which states that the use must be for the purpose of research or study, criticism or review, parody or satire, reporting news, judicial proceedings or professional advice by a lawyer, patent attorney or trade mark attorney. Then, it comes to the question of whether such use would be considered “fair”. According to Denning LJ in the landmark case *Hubbard v Vosper*:³¹

You must consider first the number and extent of the quotations and extracts. Are they altogether too long to be considered ‘fair’? Then you must consider the use made of them. If they are used as a basis for comment, criticism or review, that may be ‘fair dealing’. If they are used to convey the same information as the author for a rival purpose, that may be unfair. Next you must consider the proportions. To take long extracts and short comments may be unfair. But short extracts and long comments may be fair. Other considerations may come to mind but after all, it must be a matter of impression.

19 The combination of the two steps – the narrow interpretation of the purposes in step one and the uncertainty of interpretation of fairness in step two has rendered fair dealing a less effective defence in other common law jurisdictions as compared with the US. Despite the differences, in common law jurisdictions, the application of the fair use doctrine often adopts an equitable rule of reason, and courts typically decide on a case-by-case basis. For instance, with respect to purpose and character of use, “a finding of transformative use strongly weighs in favour of fair use, while a finding that the use is commercial in nature weighs against it. Further, the more transformative a work is, the less significant the work’s commercial nature will be to the court’s analysis”.³²

20 Contracts may limit the fair use doctrine, requiring that any use that may constitute fair use still requires permission from the right holders. This has gone beyond the role of contracts as a complement to copyright. As it forecloses some exceptions that are otherwise available, it directs to a different pathway than the copyright law (see Part IV.A).

21 In addition to the scenario-based analysis, official positions have been proposed to enhance the complementary role of contracts in relation to AI regulation in general. For instance, the opinions issued by the High People’s Court in Guangdong Province, China, provide that:³³

30 Copyright Act 1968 (Cth) ss 40–43.

31 *Hubbard v Vosper* [1972] 2 QB 84 at 94.

32 Jessica L Gillotte, “Copyright Infringement in AI-Generated Artworks” (2019) 53 *UC Davis Law Review* 2655 at 2681.

33 Zheng Caixiong & Li Wenfang “Guangdong First to Issue Judicial Policy on IP Disputes” *China Daily Hong Kong* (29 April 2025) <<https://www.chinadailyhk.com/hk/article/610640>> (accessed 11 September 2025).

Courts [in the province] should properly handle disputes over the ownership of AI-generated content and accurately define the nature of the generated content and its ownership ... The ownership of the rights or interests of AI-generated content shall be determined in accordance with the law, taking into consideration the actual agreement, the substantial contributions of each party, intellectual input, and other factors.

22 For another instance, there have been proposals that seek to ensure that “the UK’s legal framework for AI and copyright supports the UK creative industries and AI sector together”.³⁴ Contracts have been mentioned in a separate section. It recognises that control over how rights are reserved often lies not with the original creator or performer but with the platforms. Consequently, many creators in these situations feel they lack real control, as standard service contracts usually impose broad licences with little choice. Such terms, including those allowing for AI use, may be bundled with other provisions or written in a vague manner. Creators and performers report difficulty avoiding these contracts due to industry expectations to use certain tools and platforms.³⁵ Following the conventional practice of collective management organisations, it was proposed that, to provide AI developers (especially small firms and new entrants) with easy access to licensed material, collective licenses should be made available and accessible to them. On the other hand, it was also mentioned that governments would support “good licensing practice where appropriate to ensure that right holders can seek fair remuneration for their work, and that AI developers have viable means to access datasets”.³⁶ This ultimately underpins the overall suggestion of a viable pathway forward that uses licensing agreements/contracts as a primary tool to balance the needs of both right holders and AI developers, facilitated by a text and data mining exception with rights reservation mechanisms and enhanced transparency measures.

34 UK Intellectual Property Office, *Copyright and Artificial Intelligence: Consultation* (CP 1205, 17 December 2024) <https://assets.publishing.service.gov.uk/media/6762c95e3229e84d9bbde7a3/241212_AI_and_Copyright_Consultation_print.pdf> (accessed 11 September 2025).

35 UK Intellectual Property Office, *Copyright and Artificial Intelligence: Consultation* (CP 1205, 17 December 2024) at pp 17–18 <https://assets.publishing.service.gov.uk/media/6762c95e3229e84d9bbde7a3/241212_AI_and_Copyright_Consultation_print.pdf> (accessed 11 September 2025).

36 UK Intellectual Property Office, *Copyright and Artificial Intelligence: Consultation* (CP 1205, 17 December 2024) <https://assets.publishing.service.gov.uk/media/6762c95e3229e84d9bbde7a3/241212_AI_and_Copyright_Consultation_print.pdf> (accessed 11 September 2025) at pp 19.

III. Copyright-contract conflict: illegality and pre-emption of copyright

23 The issue of copyright-contract conflict arises when there is a conflict between copyright law provisions and contract terms for the same issue, creating uncertainty over which rules prevail. It is one of the most complex issues because copyright laws vary from country to country. In addition, this is an issue about the hierarchy of law, which varies due to different configurations of legislative power at the national and state levels, and the relationship between copyright law and other laws. Examples from the US and China will be used to illustrate this complexity.

A. *The US copyright pre-emption doctrine*

24 From the US perspective, copyright is provided under the federal copyright law, which pre-empts equivalent claims made under state law, including the contract laws of the states.³⁷ The two-pronged pre-emption test has been established through a series of cases since then. Firstly, states are precluded from enforcing penalties for copyright violations if the intellectual property at issue falls within the subject matter of copyright as defined by federal law; secondly, claimed property rights are equivalent to the exclusive rights provided by federal copyright law.³⁸ There has been a fine line of cases regarding whether rights created by contracts are equivalent to those under federal copyright law, as outlined in § 106 of the Copyright Act,³⁹ and whether they are enforceable.⁴⁰ In *ProCD Inc v Zeidenberg*⁴¹ (“ProCD”), the Seventh Circuit decided that a standard-form agreement prohibiting the defendant from copying a phone directory was enforceable, even though the Supreme Court decided at that time that factual databases were not copyrightable despite significant investment made through mass labour to put the database together. The key argument is that contracts (rights

37 See Copyright Act 17 USC (US) § 301(a) (2018). It states that: “... all legal or equitable rights that are equivalent to any of the exclusive rights within the general scope of copyright as specified by section 106 ... and ... come within the subject matter of copyright as specified by sections 102 and 103 ... are governed exclusively by this title. No person is entitled to any such right or equivalent right in any such work under the common law or statutes of any State”.

38 US Department of Justice, *Criminal Resource Manual No 1844: Copyright Law – Pre-emption of State Law* (archived 19 October 2020) <<https://www.justice.gov/archives/jm/criminal-resource-manual-1844-copyright-law-preemption-state-law>> (accessed 11 September 2025).

39 Copyright Act 17 USC (US).

40 Guy A Rub, “Copyright Survives: Rethinking the Copyright-Contract Conflict” (2017) 103 *Virginia Law Review* 1141.

41 86 F 3d 1447 (7th Cir, 1996).

in personam) and copyright (rights *in rem*) are different in nature, therefore rights created by contracts are not equivalent to those created by copyright law. Under such circumstances, the contracts were found to be enforceable since the second requirement of the pre-emption test was not satisfied.

25 This approach has been criticised by US commentators as it would, in particular, negatively impact the doctrine of fair use: the logic of this decision would allow right holders to amplify their statutory rights by simply removing copyright exceptions *via* contracts. Some commentators argue that the *ProCD* approach would lead to the death of copyright as it contravenes one of the core policies of the Copyright Act by extending quasi-copyright protection to works that do not qualify as “original”.⁴² In practice, however, it was found that this *ProCD* approach has been followed in the last two decades by the vast majority of courts when they dealt with a pre-emption claim concerning contracts, including the Fifth, Seventh, Eleventh, Federal Circuits, and probably the Ninth, Fourth, and Eighth Circuits.⁴³

26 Nonetheless, the Second Circuit court has decided cases concerning copyright pre-emption in § 301(a) in a mechanical, formalistic way, as demonstrated in *ML Genius Holdings LLC v Google LLC*⁴⁴ (“*ML Genius*”). In this case, Genius sued Google for allegedly scraping content in the lyrics from Genius’s website and displaying song lyrics directly in Google’s search results. Genius does not own the copyrights to the lyrics, so its claim was based on Google breaching its website’s terms of service, which prohibited commercial reproduction of its content. Since Genius’s terms of service aimed to control the reproduction and distribution of the lyrics, which are core exclusive rights protected by the Federal Copyright Act, the state law claim was considered to have been pre-empted, for such a contract created property rights equivalent to the federal copyright.

27 Following that approach, if a contract prohibits one of the exclusive rights under the Copyright Act (copying, adaptation, *etc.*), it will be pre-empted by the federal copyright law without engaging with the specific context and copyright policy considerations. Therefore, it is suggested that the question should not be whether contracts

42 David Nimmer, Elliot Brown & Gary N Frischling, “The Metamorphosis of Contract into Expand” (1999) 87 *California Law Review* 17.

43 Guy A Rub, “Moving From Express Preemption to Conflict Preemption in Scrutinizing Contracts Over Copyrighted Goods” (2023) 56 *Akron Law Review* 303 at 308.

44 No 20-3113 (10 March 2022, 2d Cir) (US) (Summary Order).

are equivalent to copyright, but whether and when contracts might undermine federal copyright policy.⁴⁵

28 The decision in *ML Genius* has significant implications for data and text mining in the context of AI training. This is because the case directly involves using contracts as a defence for data scraping. As reported, other popular websites, such as Facebook, LinkedIn, American Airlines, Ticketmaster, and even Craigslist, sued companies that scraped and copied their posted information for breach of their standard form contracts.⁴⁶ Despite the different interests that those contracts aim to protect, the formalistic approach taken by court in *ML Genius* was criticised as it failed to differentiate between contexts and focused solely on whether the enforcement of an agreement based on state law should be pre-empted by federal copyright law. However, it was also noted that this was only the reasoning of the Second Circuit, and there is a clear divergence of approaches among other US federal circuit courts.

29 The US doctrine of copyright pre-emption shows that the intricacy of the copyright-contracts relationship is embedded in the legal architecture of a country. In the US, this is reflected in the interaction between federal and state law. However, the doctrine only answers the question of which rules will prevail in the case of conflict, and nothing suggests the substantive effect on the scope of copyright protection. Indeed, that effect will entirely depend on what the contract says. As will be discussed in Part IV.A, contracts can either expand copyright protection or remove copyright exceptions.

30 Another issue that is not really clear about the copyright pre-emption doctrine in the US is whether a contractual right that prevents the other party from using data for the purpose of AI training can be considered “equivalent” to copyright. For instance, X’s *Developer Agreement and Policy* (4 June 2025) states that: “You shall not and you shall not attempt to (or allow others to): ... (k) use the X API or X Content to fine-tune or train a foundation or frontier model.”⁴⁷

45 Guy A Rub, “Moving From Express Preemption to Conflict Preemption in Scrutinizing Contracts Over Copyrighted Goods” (2023) 56 *Akron Law Review* 303.

46 Eric Goldman, “Contractual Control Over Information Goods After *ML Genius v Google* (Guest Blog Post)”, *Technology & Marketing Law Blog* (18 July 2023) <<https://blog.ericgoldman.org/archives/2023/07/contractual-control-over-information-goods-after-ml-genius-v-google-guest-blog-post.htm>> (accessed 11 September 2025).

47 X Corp, “Developer Agreement and Policy” (2023) <<https://developer.x.com/en/developer-terms/agreement-and-policy>> (accessed 11 September 2025).

31 The question is whether the right to prohibit the other party from using X content for the said AI model training is equivalent to copyright, in particular when some of the content may not be copyrightable in the first place.⁴⁸ If they are not equivalent to copyright, a *de facto* right *in personam* may lead to the conclusion that contracts can create a new right similar to copyright. If they are considered equivalent to copyright, they will be subject to the pre-emption doctrine. The *de facto* holder of the content would have to face a dilemma here. If platforms like X want their contract provisions to be effective (not pre-empted by the pre-emption doctrine), they would have to argue that the right is not an exclusive/property right like copyright.

B. *The Chinese legal principle legislative mandates pre-emption*

32 In the Chinese context, there is a long-standing recognition that intellectual property rights are statutory rights that were influenced by the civil law tradition.⁴⁹ Under this conceptualisation, copyright is a right against the world (rights *in rem*) while contractual rights are rights against a specific person (rights *in personam*). There is no explicit copyright pre-emption doctrine in China, but the pre-emption of copyright can be a derivative of the principle that intellectual property should be established by legislation.⁵⁰ This general principle only addresses the priority of copyright in the stage of right creation; nothing is said about the exercise of rights, *eg*, whether contracts can override exceptions available in copyright law. Nonetheless, as a civil law jurisdiction, interactions between contracts and copyright law in China are more complex, as one needs to take into consideration the overarching provisions in the Civil Code of the People's Republic of China⁵¹ ("Chinese Civil Code") and other legislation that aims to complement intellectual property law, such as anti-unfair competition law. These issues will be tangentially mentioned if relevant.

33 There is a general legal principle which states that contracts may not contravene laws, regulations, public order, and good morals, which consists of multiple dimensions. Firstly, Art 153(1) of the Chinese Civil Code provides: "A civil juristic act that violates the

48 Guy A Rub, "Moving From Express Preemption to Conflict Preemption in Scrutinizing Contracts Over Copyrighted Goods" (2023) 56 *Akron Law Review* 303.

49 See detailed discussion on the difference between rights *in personam* and rights *in rem* in Part IV.B.

50 Yi Jiming (易继明), "Statutory Doctrine of Intellectual Property and Its Flexibility: An Analysis of Article 123 of the General Principles of Civil Law" (知识产权法定主义及其缓和 - 兼对《民法总则》第123条条文的分析) (2017) 5 *Intellectual Property* (知识产权) 3.

51 (effective 1 January 2021).

mandatory provisions of laws or administrative regulations shall be invalid. However, this shall not apply if the mandatory provisions do not render the civil juristic act invalid.” Leading scholars have commented that the interpretation of the second sentence should consider the nature of mandatory provisions, combined with the purpose of the regulation.⁵² Secondly, Art 153(2) generally establishes the principle that a contract that violates public order and good morals is also void. Judicial interpretations have further confirmed this. In particular, the Supreme People’s Court’s interpretation on the application of the general provisions of the contract section of the Chinese Civil Code (“Judicial Interpretation of Contract Law”) provides: “even if a contract does not violate mandatory provisions of laws or administrative regulations, the people’s court shall deem it invalid under Art 153(2) of the Civil Code under any of the following circumstances . . .”⁵³ In practice, there have been consistent court cases invalidating non-compliant contracts for violating public order and good morals, particularly in areas such as leasing of dangerous buildings, nominee shareholding in listed companies, and contingency fee arrangements in criminal or administrative matters.⁵⁴

34 In general, this principle in the Chinese Civil Code, if narrowly applied to the copyright-contract relationship, means that mandatory provisions in copyright law will pre-empt any inconsistent contract terms. This copyright pre-emption is limited to copyrightable training materials. However, under Chinese law, data *per se* is not copyrightable, which means the pre-emption of copyright law is not particularly relevant in most cases related to AI training data.⁵⁵

35 Nonetheless, Chinese law recognises data as a legitimate legal interest, and copyright law is not the only legislation relevant here. China revised its Anti-Unfair Competition Law⁵⁶ (“AUCL”) in 2025. Notably,

52 Wang Liming (王利明), “Determining Contract Invalidity Based on Normative Purpose – With Focus on Article 16 of the Contract Compilation General Rules Interpretation” (论依规范目的认定合同无效 – 兼对《合同编通则解释》第16条为中心) (2024) 6 *China Legal Science* (中国法学) 24.

53 Interpretation of the Supreme People’s Court on Several Issues Concerning the Application of the General Provisions of the Contract Section of the Civil Code of the People’s Republic of China (Judicial Interpretation No 13 [2023]) (最高人民法院关于适用《中华人民共和国民法典》合同编通则若干问题的解释 (法释[2023] 13号)).

54 Rui Cai (蔡睿), “Evaluation of the Validity of Non-Compliant Contracts: Paths and Methods” (违规合同的效力评价: 路径与方法) (2025) 47 *Legal Science* 112.

55 Readers may wish to consult other contributions in this Special Issue on AI & Intellectual Property in the Asia-Pacific for complementary perspectives on this topic.

56 (2025 Revision) (effective 15 October 2025) (China). This Act was adopted at the third meeting of the Standing Committee of the Eighth National People’s Congress (*cont’d on the next page*)

it introduced a new paragraph to the Internet-related competition provision, banning unauthorised data acquisition and use. Article 13 states that:⁵⁷

... [business] operators shall not obtain or use data lawfully held by other operators by improper means such as fraud, coercion, circumvention or destruction of technical management measures, thereby damaging the legitimate rights and interests of other operators and disrupting the market competition order.

This provision is simpler than the original draft proposal, which listed four circumstances of improper means, including obtaining data by breaching agreements.⁵⁸

36 This is a significant development for companies relying on data aggregation, scraping, or competitive intelligence, and aligns with China's broader objective of clarifying the boundaries of commercial data. Despite the more general provision in the final version of the law, there has been consistent judicial practice whereby cases in AUCL are decided based on the act⁵⁹ and the legal consequences. The consequences can be referred to the previous proposal, which can be manifested through: (a) the outcome substantially substituting the main content or part of the content of network products or services lawfully provided

on 2 September 1993; amended at the 30th meeting of the Standing Committee of the 12th National People's Congress on 4 November 2017; amended at the 10th meeting of the Standing Committee of the 13th National People's Congress on 23 April 2019; and amended at the 16th Session of the Standing Committee of the 14th National People's Congress on 27 June 2025.

57 Anti-Unfair Competition Law of the People's Republic of China (2025 Revision) Art 13.

58 These four circumstances deal with:

(a) by means of theft, coercion, fraud, electronic intrusion, *etc.*, undermining technical management measures, improperly obtaining commercial data of other operators, unreasonably increasing the operating costs of other businesses, and affecting the normal operations of other businesses;

(b) obtaining and using other persons' commercial data in violation of an agreement or a reasonable and legitimate data capture agreement, which is sufficient to substantially substitute related products or services provided by other operators;

(c) disclosing, transferring or using improper means to obtain commercial data of other businesses, to an extent that is sufficient to substantially substitute the relevant products or services provided by other businesses; and

(d) improperly obtaining and using other persons' commercial data in other ways that violate the principle of *bona fide* and business ethics, seriously damaging the legitimate rights and interests of other businesses and consumers, and disrupting the order of fair competition in the market.

59 Operators using technical means such as web crawlers to obtain and use data legally held by other operators through fraud, coercion, circumvention or destruction of technical measures.

by other operators; (b) unreasonably increasing the operating costs of other operators; (c) hindering or disrupting the regular operation of network products or services legally provided by other operators and seriously damaging the rights or legitimate interests of other operators and consumers; and (d) disrupting the order of fair competition in the market.⁶⁰ The bar of proof can be high for the second and third elements listed above. However, the critical point of departure for the AUCL is that there is no stringent requirement for the aggrieved party to hold proprietary rights over the data. This is what is required: (a) the aggrieved operator legitimately held data (note they are not required to own data); and (b) the act has damaged their rights or *legitimate interests* (not the exclusive copyrights).⁶¹

37 Nonetheless, China has an administrative AI regulation titled *Interim Measures for the Administration of Generative Artificial Intelligence Services 2023*, which provides in Art 7 that:⁶²

Generative AI service providers (hereinafter referred to as providers) shall carry out training data processing activities such as pre-training and optimisation training following the law and comply with the following provisions: (a) Use data and basic models with legitimate sources; (b) Where intellectual property rights are involved, they shall not infringe upon the intellectual property rights enjoyed by others in accordance with the law....

There have not been cases concerning the validity of AI-related contracts based on Art 7. According to the use of the words “shall” and “comply with”, this article can be considered a mandatory provision although it is an administrative rule in the hierarchy of regulations. Therefore, it is possible that if a party carries out training data processing activities using data and basis models without legitimate data sources, a contract of output from these processing activities may be decided as invalid following Art 153(2) of the Chinese Civil Code and the Judicial Interpretation of Contract Law.

60 Cheng Liu (刘成) *et al*, “Key Takeaways of the 2025 Revision of China’s Anti-Unfair Competition Law” (《反不正当竞争法》2025 修订要点解读), *King & Wood Mallesons* (8 July 2025) <<https://www.kwm.com/cn/zh/insights/latest-thinking/key-takeaways-of-the-2025-revision-of-china-s-anti-unfair-competition-law.html>> (accessed 11 September 2025).

61 Anti-Unfair Competition Law (2025 Revision) (effective 15 October 2025) (China).

62 *Interim Measures for the Administration of Generative Artificial Intelligence Services* (生成式人工智能服务管理暂行办法) (effective 15 August 2023) (China). This regulatory instrument was issued by the Cyberspace Administration of China, National Development and Reform Commission, Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security and National Radio and Television Administration.

38 To summarise, under the framework of the Chinese civil law system, mandatory copyright provisions in copyright law will pre-empt conflicting contractual provisions; however, no provision explicitly prohibits contracts overriding voluntary provisions in copyright law. On the other hand, other statutes, such as the ACUL, provide a solid foundation to prohibit illegitimate access and acquisition of data, notwithstanding that these data may not be copyrightable, and these are mandatory provisions that contracts cannot override.

39 Comparatively, while there are differences between the US and Chinese approaches to dealing with the copyright-contract conflict, the basic logic behind the issues is similar – the freedom of contract manifests a liberal conception, but it only exists to the extent that the law allows. In the US, the copyright pre-emption doctrine follows the logic of legislative pre-emption based on the hierarchy of laws. In China, the legislative pre-emption further raises the question as to whether the provision in law is mandatory. Only mandatory legislative provisions prevail. If the law itself is a voluntary provision, then the parties' intention manifested in contract terms prevails.

C. *Copyright-contract conflict and doctrine of illegality*

40 The copyright-contract conflict, at a more abstract level, is governed by the doctrine of illegality in contract law. If a contract, either in its formation or performance, involves conduct that is expressly or impliedly prohibited either by common law or legislation, such illegality may lead to the effect that the contract becomes void and unenforceable.⁶³ Such an effect is ascribed because illegality is considered one of the vitiating factors of the contract, founded on policy-motivated external reasons rather than questioning the defective parties' personal contractual consent (such as duress, undue influence, or unconscionability). In addition to directly conflicting with legislation, a contract may also be deemed illegal and unenforceable on the basis of public policy considerations. This is not only a common law doctrine, but also reflected in the above discussion of the Chinese Civil Code.

41 Admittedly, illegality is a much obscure and complex area in contract law. It reflects broadly on how contracts, as an instrument of private ordering, interact with state law (primarily copyright law in this case), in particular, where the boundaries of freedom of contract

63 Australian Copyright Law Review Committee, *Copyright and Digital Agenda: Report on the Effectiveness of the Digital Agenda Amendments* (Report No 17, October 2002) at p 173 <<https://www7.austlii.edu.au/au/other/clrc/2002/17.pdf>> (accessed 11 September 2025).

should be set when statutory duties are involved. The US and Chinese laws are illustrative of typical approaches taken by different jurisdictions. The US courts, in developing the pre-emption doctrine, primarily took a mechanical approach to reserve the federal legislative power by guaranteeing that contract terms do not derogate from federal statutory copyright law. The Chinese approach focuses more on the purpose of the regulation, asking whether the exercise of contractual rights would, in effect, undermine the statutory mandate.

42 These differences can lead to nuances, especially regarding copyright exceptions.⁶⁴ Chinese law may not prohibit contract override in the sense that copyright exceptions are not considered rights or mandates within copyright law. In contrast, US law strongly suggests that contractual terms removing exceptions like fair use would be automatically invalid and unenforceable.⁶⁵ Australian law adopts a more explicit stance on contract overrides – s 47H of the Copyright Act⁶⁶ clearly prohibits contracts from excluding certain copyright exceptions, although the law remains silent on other exceptions.⁶⁷

43 Freedom of contract only exists where the law allows. Despite the differences, it is clear that, according to both the doctrine of illegality and the US doctrine of copyright pre-emption, private ordering through contracts is secondary, and freedom of contract can only be exercised within the scope defined by copyright law. However, this is based on the assumption that relevant copyright legislation exists in the first place, which is not the case for various issues concerning AI training. Consequently, the issue of whether contracts can legitimately regulate data that are not copyrightable in the first place remains, as these may be considered issues beyond the doctrine of illegality, which renders the necessity to discuss contract override in much more detail.⁶⁸ On the other hand, it should be recognised that uncertainty remains – as soon as copyright law clarifies the issue, contracts may no longer override, though these depend on the nature of copyright provisions (mandatory or not) and their retroactivity.

64 See Part IV.A specifically regarding data used in AI training.

65 Viva R Moffat, “Super-Copyright: Contracts, Preemption, and the Structure of Copyright Policymaking” (2007) 41 *UC Davis Law Review* 45.

66 Copyright Act 1968 (Cth).

67 See Part IV.A.

68 See Part III.C.

IV. Emerging contract override for regulating AI training data

44 The issue of contract override has been discussed in the literature. In particular, the 2010 UK IPO Report defines contract override as the phenomenon that occurs when contract terms supersede copyright exceptions and limits. It also pointed out that whether a contract can vary copyright exceptions is decided on a case-by-case basis.⁶⁹

45 Contract override can be justified by the liberal conception of contract in which the freedom of contract is an important legal principle. However, the exceptions under fair use (*eg* parody, citation, private copying, criticism, and news reporting) are essential to safeguard public interests and freedom of speech, which are fundamental to a democratic society. Therefore, there is a strong argument that an individual user should not be forced (which is often the case, as they do not have the bargaining power to change standard form contracts) by contract to relinquish their freedom of speech or their privacy, as contracts cannot override public policy norms.⁷⁰

46 The challenge lies in reconciling conflicting objectives – specifically on what grounds contract override can be pre-empted by state regulation. Traditional copyright exceptions are not generally considered rights. For instance, a European scholar argued that copyright limitations and exceptions are not rights but rather interests or liberties, or a “claim to the application of a rule of objective right”.⁷¹ This ambiguity makes them vulnerable when it comes to the question of whether contracts can remove such interests if they are not constructed as a legal mandate. Therefore, there is a call for granting clear legal status to copyright exceptions as rights,⁷² and an extensive interpretation of them as part of fundamental rights.⁷³

47 This has led to a divided recognition of the status of copyright exceptions. In the case that exceptions are not mandated by law,

69 Martin Kretschmer *et al*, *The Relationship Between Copyright and Contract Law* (1 July 2010) Intellectual Property Office (UK) Research Paper No 2010 (04).

70 Martin Kretschmer *et al*, *The Relationship Between Copyright and Contract Law* (1 July 2010) Intellectual Property Office (UK) Research Paper No 2010 (04).

71 Martin Kretschmer *et al*, *The Relationship Between Copyright and Contract Law* (1 July 2010) Intellectual Property Office (UK) Research Paper No 2010 (04) at p 84.

72 F Willem Grosheide, “Copyright Law from a User’s Perspective: Access Rights for Users” (2001) 23 *European Intellectual Property Review* 321.

73 Thomas Hoeren, “Copyright Dilemma: Access Right As a Postmodern Symbol of Copyright Deconstruction?” in *Digital Rights Management: Technological, Economic, Legal and Political Aspects* (Eberhard Becker *et al* eds) (Springer, 2003).

parties can therefore have the freedom of contract around it unless the law prohibits it. Some jurisdictions, such as Belgium, Ireland, and Portugal, have legislation preventing contract override, rendering such terms void.⁷⁴ The UK also amended its legislation regarding copyright exceptions for research and education in 2014.

48 This evolution of contract override in the literature – *ie* contracts removing copyright exceptions – crystallised an important aspect of the copyright-contract relationship. However, this traditional conception of contract override may not cover all circumstances where contract terms are prioritised when they conflict or have the potential to conflict with copyright law. For instance, the interaction between copyright and contracts concerning exceptions – such as fair use – does not necessarily eliminate the exceptions. Traditional contract override may not apply to circumstances where contractual terms extend copyright exceptions. It does not cover the issue of whether and to what extent contracts may intervene when the training materials are not themselves copyrightable. In addition, contract override may occur when multiple contracts co-ordinate to achieve the overriding objectives.

49 Therefore, this section proposes a broader *concept* of contract override that encompasses any circumstances in which contracts are prioritised when they are inconsistent with copyright law. It is noted that in the US, particularly when commentators discuss the case of *X Corp v Bright Data*,⁷⁵ they frame this issue as contract pre-emption – *ie* whether, in the circumstances of inconsistency, the contract undermines the policy objectives underlying US copyright law. If so, such a contract should not be enforceable as a matter of federal law, based on the conflict pre-emption doctrine.⁷⁶ The authors find that this is merely a difference in terminology and will use “contract override” throughout to cover such circumstances.

50 In addition, because copyright law is also struggling to identify its objectives when it comes to AI regulation – and there is a lack of empirical data supporting the argument of “undermining” – the authors will not

74 Martin Kretschmer *et al*, *The Relationship Between Copyright and Contract Law* (1 July 2010) Intellectual Property Office (UK) Research Paper No 2010 (04).

75 No 3:23-cv-03698 (9 May 2024, ND Cal) (US).

76 There were a series of similar cases where a platform sues a web scraper for breaching the Terms of Service, and the platform loses. The said case was decided on the basis of the conflict pre-emption doctrine, dealing with the state-federal law relationship in the US. See Eric Goldman, “X Corp v Bright Data is the Decision We’ve Been Waiting For”, *Technology & Marketing Law Blog* (17 May 2024) <<https://blog.ericgoldman.org/archives/2024/05/x-corp-v-bright-data-is-the-decision-weve-been-waiting-for-guest-blog-post.htm>> (accessed 11 September 2025).

further explore whether the effects of the contract would undermine copyright policy objectives, as this depends on which copyright objectives are emphasised in a specific context. Instead, the authors will identify the circumstances in which contracts are prioritised in the case of inconsistency in the first place.

A. *Copyright-contract relationship concerning fair use/fair dealing and other exceptions*

51 Fair use or fair dealing is an exception to copyright protection and a widely used defense against copyright infringement claims. Recent cases in the US have relied heavily on the fair use doctrine to resolve copyright issues involving training data. However, the role that contracts play here is still uncertain, as the situation is evolving and the copyright status of data is ambiguous. Nonetheless, the copyright-contract dynamics concerning copyright exceptions can be observed in the case of using materials for educational purposes by academic institutions.

52 The structure of copyright exceptions varies across jurisdictions. Australian law is selected here as an illustration for materials used for educational purposes. In the Australian Copyright Act 1968,⁷⁷ universities are provided with educational statutory licensing scheme provisions which allow multiple copies to be made by or on behalf of an academic institution for educational purposes, including the reproduction and/or communication of works in electronic form. If this statutory licensing is broadly considered as part of Australian copyright law, then there are two scenarios involving the intersection of copyright law and contracts. First, if copying (by academics and students) does not fall under one of the explicitly allowed circumstances in the statutory licensing scheme, two possible defences can be made. One is fair dealing, and the other is relying on a contract provision concerning access to electronic resources that explicitly permits copying or reproduction – should such a provision have been incorporated into the contract in the first place, of course.⁷⁸ Here, in the second line of defence, the contract plays a role in expanding the scope of copyright exceptions that are not otherwise explicitly provided for under the copyright law.

53 The second scenario is more complex. It arises when the said copying cannot be categorised as one of the explicitly allowed circumstances in the statutory licensing scheme and the user relies on a

77 (Cth) Parts VA and VB.

78 Marita Shelly & Margaret Jackson, “Copyright and Contracts: the Use of Electronic Resources Provided by University Libraries” (2012) 12 *Legal Information Management* 124.

fair use defence; meanwhile, there is a contract provision that does not allow copyright exceptions, including those that may first be introduced in the statutory licensing scheme and those offered under the fair dealing doctrine. This is the situation where traditional contract override takes charge. If the contracts are properly formed, they will be enforceable, and the copyright exceptions will be nullified.

54 It is difficult to verify the extent to which such contract override exists in reality and to estimate their impacts because copyright contracts are not often publicly available. A study of 12 contracts between publishers and research institutions in Australia found that only two permitted explicit copyright exceptions for multiple reproduction or communication under Part VB of the Copyright Act, and only three permitted fair dealing under the Copyright Act 1968.⁷⁹ This small example does prove the existence of traditional contract override.

55 One can draw some observations on the relationship between copyright and contracts regarding fair use/fair dealing from this academic publishing context, which may have implications for using data for AI model training. First, a contract can either affirm or expand exceptions, or it can eliminate the exceptions, such as fair dealing. This means that contracts interact bidirectionally with copyright protection – they can either strengthen or limit the scope of copyright protection, depending on the contract terms. Traditional contract override only covers contracts, strengthening copyright protections by removing exceptions. It has been observed that fair use has been extensively used as a defence in the existing copyright infringement cases in the US.

56 However, contracts also provide opportunities for rights holders to choose only to have “some rights reserved”, rather than “all rights reserved”. This has been the case in the Creative Commons initiative and the free software movement, which aims to promote collaboration among artists and computer programmers by rejecting the proprietary ownership of their works underpinned by the copyright system. Indeed, open source has been an option for AI model training. For instance, it is reported that under the Institutional Data Initiative (“IDI”), Harvard Law School Library has released nearly one million digitised public domain works to improve data used for AI training. The collection comprises works written in 254 languages, dating back to the 15th century.⁸⁰ The

79 Marita Shelly & Margaret Jackson, “Copyright and Contracts: the Use of Electronic Resources Provided by University Libraries” (2012) 12 *Legal Information Management* 124.

80 Harvard Law School Library, “Food for AI Thought: The Library Initiative Improving AI’s Digital Diet”, *Harvard Law Today* (25 April 2024) <<https://hls.harvard.edu/>> (cont’d on the next page)

purpose of this initiative is to ensure AI has open access to a diet of better data and promote collaboration between libraries and commercial players in the field. By training on extensive open datasets from libraries, AI absorbs better information, which also greatly enhances the technology's ability to distinguish between competing ideas. The Library and Information Services at Harvard Law School considers it a rare opportunity for altruism and business interests to be in lockstep: "Better functioning AI is better for business, and better for the world."⁸¹

57 With the expansion of institutions joining the IDI, this model can further develop as a framework for open access to AI training data. When opening the webpage of the dataset, it is unsurprising to find that contracts govern access to the system. The first clause in the Terms of Use provides that the service is for "non-commercial use only".⁸² It is also made explicit that anyone affiliated with a commercial organisation or planning to use the service for commercial purposes (including AI model training) needs to contact the owner of the data separately.

58 The second observation is that, even within the same category of licensing agreements with similar parties, the permitted uses vary significantly, depending on the specific contract terms governing the issue. For instance, contracts can impact the uses in terms of what is exactly permitted, as shown in the example of licensing agreements between publishers and universities.⁸³ In another instance, a derivative work based on one or more pre-existing works in some way, such as a translation, adaptation, or sequel, can be considered a new work, despite retaining some of the characteristics of the original work.⁸⁴ The current judicial jurisprudence in the US is that if a work can serve a transformative purpose, it is itself copyrightable even if it does not alter the content of the original work.

today/food-for-ai-thought-and-the-library-initiative-improving-ais-digital-diet/> (accessed 11 September 2025).

81 Harvard Law School Library, "Food for AI Thought: The Library Initiative Improving AI's Digital Diet", *Harvard Law Today* (25 April 2024) <<https://hls.harvard.edu/today/food-for-ai-thought-and-the-library-initiative-improving-ais-digital-diet/>> (accessed 11 September 2025).

82 Clause 1 of the Terms of Use for Early-Access provides that: "You may use the Service solely for non-commercial purposes. Open-source projects and other public-use efforts are welcome, even if they may indirectly support commercial use, so long as they are unaffiliated with commercial actors or intent."

83 Marita Shelly & Margaret Jackson, "Copyright and Contracts: the Use of Electronic Resources Provided by University Libraries" (2012) 12 *Legal Information Management* 124.

84 R Anthony Reese, "Transformativeness and the Derivative Work Right" (2008) 31 *Columbia Journal of Law & the Arts* 467.

59 However, some categorical interventions by statutes may alter the outcome of competition between contracts and copyright law. It is clear that allowing contracts to waive copyright exceptions may jeopardise other objectives in copyright law, in particular, the public interest in access to knowledge. In certain jurisdictions, contract override for educational purposes has been further regulated by the state. For instance, in the UK, certain copyright regulations state: “To the extent that a term of a contract purports to prevent or restrict the doing of any act which, by virtue of this section, would not infringe copyright, that term is unenforceable.”⁸⁵ As a result, library, research, and education copyright exceptions in the UK cannot be overridden by a contract term purporting to remove such exceptions.

60 Similar concerns exist in the US. As commented by Moffat:⁸⁶

[B]y permitting copyright owners to contract around fair use, courts have improperly abdicated their fair use policymaking role while at the same time arrogating to themselves policymaking regarding contracting around fair use, which is a task that should be placed at Congress’s door.

This shows that in the US, contract override is not only an issue between copyright law and contracts, but also the legislative power of Congress to amend the Copyright Act and the judge-made law permitting the extent to which a contract can remove copyright exceptions.

61 The European Union’s (“EU”) text and data mining exception, as outlined in Arts 3 and 4 of the 2019 Copyright in the Digital Single Market Directive,⁸⁷ has provided an interesting example of statutory affirmation of the exceptions.⁸⁸ For instance, it requires that “Member

85 This provision is found in three regulatory instruments in the UK: (a) The Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014 (SI 2014/1372) (UK) (entered into force 1 June 2014) <<https://www.legislation.gov.uk/ukxi/2014/1372/contents/made>> (accessed 11 September 2025); (b) The Copyright and Rights in Performances (Disability) Regulations 2014 (SI 2014/1384) (UK) (entered into force 1 June 2014) <<https://www.legislation.gov.uk/ukxi/2014/1384/contents/made>> (accessed 11 September 2025); and (c) The Copyright and Rights in Performances (Quotation and Parody) Regulations 2014 (SI 2014/2356) (UK) (entered into force 1 October 2014) <<https://www.legislation.gov.uk/ukxi/2014/2356/contents/made>> (accessed 11 September 2025).

86 Viva R Moffat, “Super-Copyright: Contracts, Preemption, and the Structure of Copyright Policymaking” (2007) 41 *UC Davis Law Review* 45.

87 Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market [2019] OJ L 130/92 (“CDSM”).

88 The Art 3 CDSM exception for text and data mining regulates “research organisations and cultural heritage institutions ... for the purpose of scientific research”. Article 4 provides a more qualified exception but does not restrict the entities and purposes
(*cont’d on the next page*)

States shall provide for an exception or limitation to the rights” provided under existing EU directives. This means the exceptions are not optional but mandatory. It resembles the UK 2014 copyright exceptions discussed above. This means contracts cannot override (remove) such an exception by requiring comprehensive copyright protections under relevant directives. However, in Art 4, two conditions were stated. Firstly, the person who purports to reproduce or extrude text and data must have lawful access. The first condition is consistent with the US ruling in *Bartz*. Secondly, the works or other subject matter have not been expressly reserved by their rights holders in an appropriate manner, such as through machine-readable means in the case of content made publicly available online. This second condition (rights holder’s opt-out) is critical regarding contract overridability – it can be read as an exception to the text and data mining exception. Therefore, if there is an agreement on data to be used for AI training, the order of priority is as follows: (a) rights holder’s claim to opt out where the rights holder is a third party to the platform-AI developer agreement, provided lawful access is granted *via* a licensing agreement; (b) statutory affirmation text and data mining exception which contract terms may not override; and (c) other aspects concerning the reproduction and extraction of text and data.

62 Given the variety of situations in which contracts may override copyright, there is no definitive answer as to how this issue may evolve concerning AI training data. Private actors, such as Harvard Law School Library, have taken proactive steps by permitting non-commercial use of AI training data, although legislation remains unclear on this point. No matter which direction is taken, the general principle remains – the question of overridability must be determined in a way that reflects both the private and individual interests and wishes of the parties, as well as the public interest in maintaining the overall balance on which the legislation is based.⁸⁹

B. Contracts as a rights-creation mechanism

63 As discussed in the section above, contracts not only override copyright in its traditional sense of removing copyright exceptions, but they may also be used to create a new *de facto* right that is similar to copyright, based on the contractual promise. In this case, there are some challenges regarding contract override – the term may be used

(including commercial use), but such an exception is subject to being opted-out by right holders.

89 Thomas Riis & Jens Schovsbo, “How Much ‘Freedom of Contract’ in EU Copyright Law?” in *Kreation Innovation Märkte-Creation Innovation Markets: Festschrift Reto M Hilty* (Florent Thouvenin *et al* eds) (Springer, 2024).

as a misnomer. At the time of forming such a contract, if there were no copyright protection, there would be nothing for the contract to override. Prospectively, suppose that copyright law expands its subject matter in certain aspects concerning AI training, then copyright law will prevail, as conflicting contract provisions will be deemed illegal. For instance, after OpenAI's launch of GPT-4o, a new AI-powered feature allowed users to generate high-quality images by transforming internet memes and personal photos into artwork resembling the distinct style of Hayao Miyazaki, the founder of Studio Ghibli.⁹⁰ Imagine that in the Ghibli studio scenario, where style is not currently protected under the copyright law, a contract that prevents using AI to generate outputs of a similar style may create a right that converges with the objective of the copyright law to protect original creators. It does not conform to the contract overriding scenario discussed before. There is nothing that a contract can override; it may be the only regulatory tool available for the *de facto* rights holder. If, however, copyright law reaffirms its position in the idea-expression dichotomy and states clearly that styles are not copyrightable specifically concerning AI training, then any contract provision will be unenforceable after such a law is effective.

64 There is a question of whether contracts, as a private regulation mechanism, can create new rights in these cases that are not available under copyright law. As contracts have been so ubiquitously used, it has been argued that instead of a regulatory tool for private ordering, contracts work more like private legislation.⁹¹ Nonetheless, it has been argued that copyright law reflects a delicate balance between an owner's monopoly and a user's privilege to access information; a balance that may be ignored or neglected by such private legislation. Copyright endowments have also been considered a result of a social contract, where individuals within society reach a mutually beneficial agreement to be governed by a state authority.⁹² Any attempt to alter the balance struck by the copyright law privately may distort the federal copyright law in the US. In other jurisdictions, this may raise similar concerns of illegality.⁹³

90 Swagita Pandey Tewari, "Ghibli AI Art and Copyright: The Copyright Dilemma of Ghibli Style Creations and the Looming Legal Battles Ahead", *The IP Press* (31 March 2025) <<https://www.theipress.com/2025/03/31/ghibli-ai-art-and-copyright-the-copyright-dilemma-of-ghibli-style-creations-and-the-looming-legal-battles-ahead>> (accessed 11 September 2025).

91 Viva R Moffat, "Super-Copyright: Contracts, Preemption, and the Structure of Copyright Policymaking" (2007) 41 *UC Davis Law Review* 45 at 69

92 Alina Ng, "The Social Contract and Authorship: Allocating Entitlements in the Copyright System" (2008) 19 *Fordham Intellectual Property Media & Entertainment Law Journal* 413.

93 See Part III.C.

65 However, the question here is different. As copyright law defines the initial endowment in original works of authorship,⁹⁴ arguably, the prerequisite for copyright pre-emption to work is that the relevant subject matter is original and can be protected under copyright law, and the exclusive rights are themselves already recognised as copyright. If the copyrightability of AI training data is unclear, the foundation for copyright pre-emption does not exist. There can be an argument that it is not an issue of contracting around copyright. The rights created by contracts in this case are not relevant to copyright, and copyright may not pre-empt them anyway because copyright law is silent on the issue outside of its remit. Therefore, nothing explicit in statutory law or legal doctrine prevents the creation of new rights through contracts that limit or prevent the use of data over which they have *de facto* control for AI training.

66 However, it is still important to understand the difference between copyright law and contracts. Rights created by contracts are rights *in personam* (rights directed to a person, who is the other party of the contract), and copyright, as a statutory right, is a right *in rem* (also known as an absolute right,⁹⁵ which is a right against the world). The distinction between *rights in personam* and *rights in rem*, in particular, is a special notion in the continental system of private law. Some argue that they are not translatable into the common law language.⁹⁶

67 What is interesting here is the difference in function. When a right is created through a contract, *eg*, the right to prevent users from using data from a platform for AI training, the cause of action is a breach of contract instead of copyright infringement. The term that is breached is an explicit prohibition on using their data for training AI models, as stated in the terms and conditions for service on their website.⁹⁷ However,

94 Niva Elkin-Koren, "Copyright Policy and the Limits of Freedom of Contract" (1997) 12 *Berkeley Technology Law Journal* 93 at p 98.

95 In absolute legal relationships, the entitled party is certain; therefore, the other party on the opposite side is generally an uncertain group of people. By contrast, legal relationships that arise from obligations – such as contracts, unjust enrichment, or torts – are usually relative. See József Benke "Fundamentals of Law of Rights In Rem" in *The Fundamentals of Hungarian Private Law* (József Benke ed) (University of Pécs Faculty of Law, 2020) ch 5 at p 28.

96 József Benke "Fundamentals of Law of Rights In Rem" in *The Fundamentals of Hungarian Private Law* (József Benke ed) (University of Pécs Faculty of Law, 2020) ch 5.

97 Stuart D Levi, Mana Ghaemmaghami & MacKinzie M Neal, "Online Terms of Use and the Training of AI Models", *AInsights* (Skadden, July 2023) <https://www.skadden.com/-/media/files/publications/2023/07/online_terms_of_use_and_the_training_of_ai_models.pdf> (accessed 11 September 2025).

such a claim cannot be made against anyone; it can only be made against the party that has presumptively accepted the terms of the contract.

68 Therefore, the key difference is that the rights created by contracts (not using data for training purposes) are directed to the other party and are not, in nature, equivalent to the exclusive copyright against the world at large. Critically, it is up to the other party's assent for such a term to be enforceable. In general, a term of a contract can be incorporated by signature, and the objective approach does not require the other party to read and understand the content of the terms.⁹⁸ In the online environment, if clicks can be proven through clickwrap agreements or scrollwrap agreements, it is presumed that assent has been given and the contracts are generally held enforceable.⁹⁹

69 In the case of unusual and unexpected terms (*ie*, terms that include particularly onerous obligations) in a standard form contract, common law requires that the party seeking to incorporate the terms must make extra efforts to ensure that the terms are incorporated.¹⁰⁰ Proving assent is more challenging for browsewrap agreements than clickwrap and scrollwrap agreements. As actual notice is often absent in the online environment, the real question turns into one about what sort of website design elements would be necessary or sufficient to deem a browsewrap agreement valid. And it is confirmed by *Breman v Freedom Financial Network*¹⁰¹ that a browsewrap agreement is often unenforceable *per se*. Nonetheless, in an online environment, the validity of a browsewrap agreement may eventually turn on what sort of website design would draw sufficient attention from a reasonably prudent user. On the other hand, court decisions on browsewrap agreements may not be relevant anymore as soon as such contracts evolve or devolve, which makes it more difficult to categorise a typical browsewrap contract.¹⁰²

98 Typical common law cases include *L'Estrange v F Graucob Ltd* [1934] 2 KB 394 and *Toll (FGCT) Pty Ltd v Alphapharm Pty Ltd* [2004] HCA 52; 219 ALR 13.

99 In clickwrap agreements, website users are required to click on an "I agree" box after being presented with a list of terms and conditions of use. In scrollwrap agreements, the terms and conditions are placed inside a scrollable window on the screen, users have to physically scroll down to pass through the entire agreement before accepting, giving the website stronger evidence that the user had an opportunity to read it. See *Sellers v JustAnswer LLC*, 289 Cal Rptr 3d 1 (Ct App, 2021). See also Nathan J Davis, "Presumed Assent: The Judicial Acceptance of Clickwrap Part II: Cyberlaw: Section A: Notes" (2007) 22 *Berkeley Technology Law Journal* 577.

100 See *Interfoto Picture Library Ltd v Stiletto Visual Programmes Ltd* [1989] QB 433.

101 30 F 4th 849 at 868 (9th Cir, 2022).

102 Sophie Lee, "Getting a 'Bad Wrap': An Analysis of Online Contract Cases in California after Step-Saver and ProCD" (2024) 98 *Southern California Law Review* 419.

70 In the case of incorporation by notice, what is also relevant is the designated way of acceptance. It is not a difficult issue for the content provider – they could simply define the use of the product/service as a method of acceptance. In the case of a unilateral contract, anyone who uses the service becomes a party to the agreement.

71 If a term “No AI training. Provider may not use customer content or usage data to train any artificial intelligence, machine learning, large language models, or other similar networks, algorithms, or systems”¹⁰³ is validly incorporated in a contract, it will bind the other party. The key question in reality is how technically difficult it may be for a “right holder” to prevent unlicensed use.¹⁰⁴ If assent can be obtained from every user, either through technology or law, there will be no unlicensed use and no user will be free from contractual restrictions. In other words, when all access to the content (even if it is not copyrightable) is controlled and permissions are granted upon agreeing to a standard form license, the terms in the contract will effectively govern all relationships between the content holder and the content user. In this situation, a contractual prohibition on using data for AI training against every individual user who assents to the contract terms will be functionally similar, if not equivalent to, an exclusive right otherwise available in copyright. Contracts can, therefore, create *de facto* rights based on the capacity to control access.

72 This was precisely the basis for Reddit to sue. Reddit’s browsewrap provision states that, “By accessing or using our Services, you agree to be bound by these Terms. If you do not agree to these Terms, you may not access or use our Services.” However, it is questionable whether and to what extent the right holder, like Reddit, can control access or use of the service. It is questionable whether data holders can have any *de facto* control at all, particularly in the case of data scraping. If they can, they would somehow technically prevent access in the first place from those who do not agree to the terms; if they cannot, they would have to argue that the other party has actually agreed to the terms. But proving assent for the browsewrap contract is difficult, in particular when the term is phrased as above.

73 In addition to defective assent, courts may also be open to making a contract voidable due to unconscionability, violations of

103 This is a sample clause provided by Common Paper, a website specialising in assisting with contract drafting. See Common Paper, “Prohibit AI Training” (Cloud Service Agreement, 2023) <<https://commonpaper.com/standards/cloud-service-agreement/prohibit-ai-training>> (accessed 23 September 2025).

104 Niva Elkin-Koren, “Copyright Policy and the Limits of Freedom of Contract” (1997) 12 *Berkeley Technology Law Journal* 93 at p 103.

public policy for consumer protection, or the fairness of forum selection clauses.¹⁰⁵

74 There is also another argument that differentiates copyright law and contracts concerning the creation of rights – copyright law defines the initial entitlement, while contracts only govern the transferability of titles that have already been determined.¹⁰⁶ This differentiation is more of an observation rather than a normative one. There are indeed intellectual property rights, such as trade secrets, that are governed primarily by private regulation in most jurisdictions. It is not that contracts may not create rights, but rather on what basis the party seeking to exercise that right can leverage enforcement. In normal intellectual property rights, such as copyright and patents, the statutory provision is the leverage. In trade secrets, *de facto* control of undisclosed information (*ie*, taking reasonable steps to maintain secrecy) is the leverage. It appears that the AI developer’s promise to refrain from using relevant data for AI training is a leverage, which nonetheless can be defective for various reasons, as illustrated above. More importantly, copyright exclusivity is justified by policy objectives. By contrast, it can be challenging for some data holders to justify their rights beyond a promise if they neither own the data nor have control over them.

75 In summary, the role of contracts in regulating AI training data varies depending on the copyrightability of the contents used for AI training: if the copyright status is clear, contracts play the same role as they do for ordinary copyright licensing agreements; if the copyright status is not clear, contracts may play the role of *de facto* rights creation. *Still*, the rights created can be fragile, due to possibly defective assent and lack of justification.

C. *Creator’s right to opt out or opt in*

76 As pointed out by the 2010 UK IPO Report, the real creators – artists and authors – have been in a powerless position as they have agreed to broad and vague licensing agreements with the platforms.¹⁰⁷ In particular, it is unclear to what extent they can control whether their works are permitted for use in AI training. Contracts have followed the doctrine of privity, where a contract only binds the parties to the

105 Niva Elkin-Koren, “Copyright Policy and the Limits of Freedom of Contract” (1997) 12 *Berkeley Technology Law Journal* 93.

106 Niva Elkin-Koren, “Copyright Policy and the Limits of Freedom of Contract” (1997) 12 *Berkeley Technology Law Journal* 93 at pp 105–106.

107 Martin Kretschmer *et al*, *The Relationship Between Copyright and Contract Law* (1 July 2010) Intellectual Property Office (UK) Research Paper No 2010 (04) at p 15.

contract.¹⁰⁸ Once the right to permit a third party to use the work is granted to a platform in a lump sum, the creators have little control over future arrangements, as they are not a party to that contract between the platform and the AI developer concerning the use of the work for AI training. The so-called opt-out model allows an author to opt out of such a deal, as the original creator of the work. It allows AI developers to train on any available data, unless a rights holder has affirmatively opted out of being included in the dataset.

77 The EU is currently developing a model of opt-out and transparency to address the two weaknesses of the 2019 Copyright in the Digital Single Market Directive – the lack of clear and recognised standards for effective reservations of rights, and the lack of transparency regarding the training data used by machine learning developers.¹⁰⁹ The opt-out by the rights holder has been used as a mechanism to clarify the boundary of reservation of rights.

78 Nonetheless, the opt-out model has been criticised for the technical challenge of implementation. The existing practice of the EU's General Data Protection Regulation¹¹⁰ with an opt-out model for privacy regulation and recent EU cases indicate the following challenges:

- (a) the use of non-standardised or unsuitable technologies;
- (b) uncertainty over the appropriate modalities for exercising an opt-out;
- (c) questions about timing (whether opting out is possible after data mining has already occurred);
- (d) the difficulty in deciding the scope of opting out, such as whether it should apply at the site level or the individual work level and the impact on downstream uses;
- (e) complex jurisdictional issues; and
- (f) the overlap with personal data protection issues.¹¹¹

108 See for instance, English case *Beswick v Beswick* [1968] AC 58 (HL) and Australian case *Coulls v Bagot's Executor and Trustee Co Ltd* (1967) 119 CLR 460.

109 Paul Keller & Zuzanna Warso, *Defining Best Practices for Opting Out of ML Training* (29 September 2023) Open Future Policy Brief No 5 <https://openfuture.eu/wp-content/uploads/2023/09/Best-practices_for_optout_ML_training.pdf> (accessed 23 September 2025).

110 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data [2016] OJ L 119/1.

111 Martin Kretschmer *et al*, "Copyright and AI in the UK: Opting-In or Opting-Out?" (2025) *GRUR International* ikaf09 at p 7.

79 The key problem with the opt-out model is that it prioritises data access over the interests of the creators. According to a UK study, the income of UK creators has dropped substantially in the last two decades. In 2022, authors earned only £7,000 from their creative work, a 60% decline since 2006 (£17,608). Similarly, in 2024, visual artists earned £12,500 annually, a 47% decrease from 2010 (£21,320).¹¹² While the decline in creators' incomes already existed before the widespread availability of AI, an alternative opt-in model was proposed in the UK as an opportunity to rebalance the interests of the parties.¹¹³ In the US, the opt-out is also criticised as “an empty promise to creators, enabling tech companies to continue their mass exploitation of unlicensed data while maintaining a veneer of compliance with IP laws”.¹¹⁴

80 Despite the criticism, Anthropic changed its terms and privacy policy for Claude in September 2025, switching from an opt-in to an opt-out model. The new terms read: “We may use your Inputs and Outputs to train our models and improve our Services, unless you opt out through your account settings. Even if you opt-out, we will use Inputs and Outputs for model improvement when: (1) your conversations are flagged for safety review ...”.¹¹⁵ It is still too early to evaluate whether this would become a trend for major AI developers.

D. How would contracts interact with transparency requirements?

81 Since AI developers frequently use content without seeking permission from rights holders, there is a requirement to disclose information on the contents of AI training data. Training data transparency requirements are a feature of the EU AI Act 2024,¹¹⁶ which is explicitly designed to support the enforcement of text and data mining under the 2019 Copyright in the Digital Single Market Directive.¹¹⁷ There is a

112 Martin Kretschmer *et al*, “Copyright and AI in the UK: Opting-In or Opting-Out?” (2025) *GRUR International* ikaf09 at p 7.

113 Martin Kretschmer *et al*, “Copyright and AI in the UK: Opting-In or Opting-Out?” (2025) *GRUR International* ikaf09.

114 Barbara Rasin, “Opt-Out Approaches to AI Training”, *Berkeley Technology Law Journal Blog* (18 April 2025) <<https://btlj.org/2025/04/opt-out-approaches-to-ai-training/>> (accessed 23 September 2025).

115 Anthropic PBC, “Privacy Policy” (28 September 2025) <<https://www.anthropic.com/legal/privacy>> (accessed 11 September 2025).

116 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence [2024] OJ L 2024/1689 (“AI Act 2024 (EU)”).

117 AI Act 2024 (EU) Art 53(1)(c). Adam Buick, “Copyright and AI Training Data – Transparency to the Rescue?” (2024) 20 *Journal of Intellectual Property Law & Practice* 182.

concern that this will result in additional costs for AI developers, which may drive them to relocate outside of the EU.¹¹⁸ However, this extreme scenario may not happen due to the role that contracts play in this process.

82 AI developers and other gatekeeper organisations may include provisions in their licensing agreements to simplify compliance with the transparency requirement. On the one hand, AI developers could cite works in training data summaries, as *per* the licensing agreement, and follow the EU copyright regulation through those terms. Other stakeholders, such as publishers, may introduce standard contract terms requiring authors to waive their right to opt out of training data mining to facilitate further such agreements.¹¹⁹ This would not be difficult to implement, given the unbalanced bargaining power of these organisations; however, it would defeat the purpose of the transparency requirement without benefiting the actual creators who were “forced” to opt out. The trickiest part of this arrangement is that it poses no problem from the perspective of contract law. It is just another example of the unfairness inherent in standard form contracts, caused by unbalanced negotiating powers.

83 Nonetheless, there is another obligation imposed on the provider of general-purpose AI models to “draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.”¹²⁰ The template was published on 24 July 2025, and all providers of general-purpose AI models have to fill it out before they enter the EU market.¹²¹ In addition to this market entry barrier, non-compliance will lead to a fine of €15m or 3% of the provider’s total worldwide annual turnover in the preceding financial year, whichever is higher.¹²² The template clearly requires disclosure of the nature of content for publicly available datasets and non-publicly available datasets obtained from third parties, and data processing aspects, which would specifically

118 Martin Senftleben, “AI Act and Author Remuneration – A Model for Other Regions?” (24 February 2024) <<http://dx.doi.org/10.2139/ssrn.4740268>> (accessed 26 October 2025).

119 Adam Buick, “Copyright and AI Training Data – Transparency to the Rescue?” (2024) 20 *Journal of Intellectual Property Law & Practice* 182.

120 AI Act 2024 (EU) Art 53(1)(d).

121 European Commission, “Explanatory Notice and Template for the Public Summary of Training Content for General-Purpose AI Models” (24 July 2025) <<https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models>> (accessed 23 September 2025).

122 AI Act 2024 (EU) Art 101(1)(a).

involve whether datasets concern personal data, copyrightable content, machine-generated data such as Internet of Things or synthetic data.

84 The EU AI Act 2024, set to be implemented from August 2026, will reveal how its mandatory public disclosure interacts with copyright infringement claims. However, as a matter of fact, the training data must be reproduced at least once as part of the training process, which is at least arguably a *prima facie* infringement of the copyright of reproduction.¹²³ If public disclosure of training content is implemented rigorously, it would bridge the missing link of the lack of evidence for copyright infringement litigation.¹²⁴

V. Conclusion

85 Contracts have been ubiquitously used in regulating AI training data. This article provides a detailed review of the role contracts play in regulating data used for AI training purposes. Regulating AI training data through contracts is like building a sandcastle at high tide – a metaphor that highlights its fragmented, fragile and temporary nature. Since private ordering through contracts is secondary to state regulation and the freedom to be exercised is subject to copyright law, the doctrine of illegality suggests that there is considerable uncertainty for contractual regulation of training data when copyright law itself is unsettled.

86 The authors found that contracts are particularly relevant when the copyright status of the training content is unclear or when it is evident that the training data are not copyrightable. As long as the contract is validly formed, breach of contract itself can be the cause of action, arguably without confirmation of copyrightability in the first place. This way, contracts create *de facto* rights, although the nature of these rights and their scope may vary depending on the contract's construction and applicable copyright law. Nonetheless, this temporal right will only survive till conflicting copyright regulation intervenes, due to the doctrine of illegality. For instance, forthcoming copyright provisions on issues of copyrightability and fair use of content for AI training may pre-empt directly conflicting contractual provisions.

123 Adam Buick, "Copyright and AI Training Data – Transparency to the Rescue?" (2024) 20 *Journal of Intellectual Property Law & Practice* 182.

124 European Commission, "Explanatory Notice and Template for the Public Summary of Training Content for General-Purpose AI Models" (24 July 2025) <<https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models>> (accessed 23 September 2025).

87 While it is generally recognised that contracts are complementary to copyright law, the authors found that contracts interact bidirectionally with copyright protection – they can either strengthen or limit the scope of copyright protection, depending on the construction of terms. In particular, it may limit the fair use doctrine, which forecloses exceptions that may be otherwise available. Recent legislative developments across the globe also indicate that debates concerning copyright exceptions have become increasingly granular, and contracts, as an integral part of the regulation, have introduced further uncertainties. Still, more empirical data is needed to understand the landscape better, including the extent to which contracts for AI training have waived copyright exceptions, which specific exceptions are involved, and how the interests of the parties have been impacted.

88 Interestingly, the opt-in and opt-out debate appears to be a contractual issue, as the authors or performers can only opt out or opt in through their original licensing agreement with the platform; however, such contract terms can be drafted either way. It is ultimately a public policy choice – whether the policymakers favour efficiency and access to train AI models to win the AI competition, or they would rather protect the original creators who have already been devastated over the last two decades.

89 The EU AI Act 2024 will clarify how mandatory public disclosure affects copyright infringement claims. Strict public disclosure of training content could fill evidence gaps in copyright cases. This transparency requirement serves as a starting point to observe how the doctrine of illegality in contract law interacts with other areas of law for issues that may have an impact on copyright.
