Proceedings of the 5th International Conference on International Law and Legal Policy
DOI: 10.54254/2753-7048/2026.36650

The Applicability of Reproduction Right During AI Model Weight Training: Comparative Research Based on Two Typical Judgments

Shenxin Yang

Shenzhen Chenghan School International Curriculum Experimental Campus, Shenzhen, China
jacksonyang.sz@outlook.com

Abstract. The fast-growing generative artificial intelligence industry have brought unprecedented challenges to existing copyright rules, especially regard how developers utilise copyrighted literary, visual and coding works to train neural network models. Legal scholars long debated one core question: does building model weights in the training process constitute a copyrighted reproduction act as define by national copyright laws and international treaties? This paper compare two landmark verdicts released in recent years: the 2024 Ultraman AI copyright dispute judged by Hangzhou Internet Court in China, and Thomson Reuters v. Ross Intelligence ruled by the District Court of Delaware in the United States in 2025. Through case analysis and comparative legal research, this paper sort out different judicial attitudes toward three core technical acts: raw data ingestion, temporary data storage in computing memory, and final weight parameter fixation. The analysis show that Chinese judges adopt an output-centred judging logic, treating temporary storage of copyrighted content during training as an inevitable auxiliary technical step without independent infringement liability. By contrast, American courts conduct a full four-factor fair use test covering every stage of AI training workflow. To balance technological progress and creators' exclusive copyright benefits, this paper put forward a two-tier "market impact balancing test" for courts to judge reproduction disputes arising from AI weight training.

Keywords: AI model training, reproduction right, neural network weights, fair use, copyright limitation and exception

1. Introduction

As generative AI tools gradually penetrate various industries, massive collections of copyrighted materials become essential raw materials for training large neural network models. The legality of copying and processing protected works in training pipelines has become a hot litigation issue cross jurisdictions. Traditional copyright legislative frameworks was drafted long before large-scale machine learning emerged, so many technical behaviours unique to AI training lack clear legal positioning [1]. Two representative judgments from China and the US offer valuable reference to resolve such emerging conflicts.

DOI: 10.54254/2753-7048/2026.36650

In 2024, Hangzhou Internet Court delivered its judgment on the dispute between Shanghai Xinchuanghua Culture and Hangzhou Jellyfish Intelligent Technology, widely known as the Ultraman AI case. This ruling stand as China's first authoritative judicial interpretation on copyright risks generated by AI training data. The defendant company trained a LoRA fine-tuning model using copyrighted Ultraman character images without formal licensing, and the model later produce images highly similar to the plaintiff's original artworks. The court classified the temporary storage of copyrighted pictures during model fine-tuning as mere temporary retention attached to feature extraction techniques, which triggered extensive discussions among domestic intellectual property researchers regard the scope of reproduction right protection [2].

One year later, the District Court of Delaware issued its judgment on Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., which presents a completely different judicial stance compare with the Ultraman case. Ross Intelligence copied Westlaw's original legal case headnotes created by editorial teams to build a competing AI legal research platform [3]. The judge thoroughly reviewed the defendant's fair use defence and refuse to recognise AI training as an automatically transformative activity, showing a restrictive attitude toward unlicensed reproduction of creative works for model training purpose.

Despite different final rulings, the two cases revolve around an identical intersection of technology and law: the legal status of the whole weight formation procedure. Weight training cover four continuous steps: importing raw copyrighted materials, forward and backward propagation cycles, parameter adjustment via gradient descent, and permanent storage of final weight matrices. The core legal puzzle remain unclear: whether loading protected works into RAM/GPU memory temporarily, or embedding original expressive content into fixed weight parameters, counts as actionable reproduction that violates copyright [4]. Furthermore, the two lawsuits adopt distinct liability identification logic: Chinese courts mainly discuss secondary liability based on platform negligence, while American judges directly examine primary infringement through the full training workflow. This paper compare judicial reasoning in both jurisdictions to clarify how the scope of reproduction right evolves under algorithmic learning scenarios.

2. Legal basis of reproduction right in AI training scenarios

2.1. Technical workflow of weight generation and legal definition of reproduction

To judge whether reproduction right applies to AI training, we first need breakdown the four practical technical phases of weight construction. First, data preprocessing convert original texts, images or audio files into tokenised tensor formats readable by machine hardware. Second, iterative forward propagation generate preliminary predictive outputs, and backward propagation distribute calculation errors across network layers to adjust operational logic. Third, gradient descent algorithms continuously revise weight values to reduce the gap between predicted results and real labels. Fourth, the finished weight matrix is saved permanently, marking the end of weight formation [5].

After clarifying technical operations, we need analyse these acts within the international copyright system. Both China and the United States are contracting parties to the Berne Convention, which list reproduction right as one of the most fundamental exclusive rights hold by authors. Reproduction generally refer to the act of fixing creative works onto tangible media without authorisation. Nevertheless, digital computing environments create new difficulty for copyright enforcement [6]. The most controversial point lie in judging whether temporary loading of works into computer memory satisfies the "fixation" requirement for identifying infringing reproduction.

DOI: 10.54254/2753-7048/2026.36650

Two opposing academic theories have formed over this debate.

The Statistical Abstraction Theory hold that model weights only record statistical correlations and probability distributions extracted from training datasets. Supporters argue that individual weight figures cannot restore any complete original work, and no recognisable creative expression exist in isolated parameters. Therefore, weight matrices cannot be treated as copies regulated by copyright law [7]. Technically, single weight values contain no retrievable information about specific training input files.

On the contrary, the Implicit Reproduction Theory focus on the "memorisation" phenomenon of large AI models [8]. When specific prompts can trigger the model to output content almost identical to original training materials, it prove that original creative expression is implicitly encoded within weight parameters [9]. Multiple security research experiments have verify that large language models can output verbatim segments from training data under certain input conditions. This mean copyrighted expressive content exists in latent form inside weight matrices, which should fall under the control scope of reproduction right [10].

The core divergence between the two theories centre on two legal standards: the legal consequence of temporary reproduction, and the threshold of fixation required to establish copyright infringement. The following sections analyse the two representative cases to compare judicial practice on these standards [11].

2.2. The Ultraman case: output-oriented judgment logic and temporary storage qualification

The court's analysis in the Ultraman case was built on Article 24 of China's Copyright Law, which adopt a closed-list system for copyright exceptions. As a civil law jurisdiction, China only permit use of copyrighted works without licensing within statutorily enumerated scenarios; judges lack authority to create new exception categories through judicial interpretation. This system differ fundamentally from the open-ended fair use mechanism applied in common law countries including the US and UK [12].

The plaintiff, the exclusive copyright holder of Ultraman visual artworks in mainland China, claimed the defendant's AI training behaviour and subsequent infringing image generation violated its reproduction and information network transmission rights. The judge separated the case reasoning into two layers [13].

First, regarding acts occurring in the training phase: the court state that temporarily storing Ultraman images during LoRA fine-tuning served only to extract visual feature data, which was an indispensable auxiliary step of technical operation. Such temporary retention could not independently constitute an infringement of reproduction right.

Second, concerning AI-generated outputs: the court compare generated pictures with original Ultraman works and confirm substantial similarity. It further rule that the platform operator failed to deploy effective content filtering, real-time monitoring and timely deletion mechanisms to block infringing outputs. Based on tort liability rules in Chinese civil law, the platform bear contributory infringement liability for neglecting its reasonable security obligations.

This output-centred judging strategy reflect deliberate policy consideration. Judges avoid making definitive rulings on whether weight matrices count as copies by shift infringement judgment to the final output stage. This method bring flexibility to judicial practice, as courts can directly address tangible market harm without getting trapped in obscure technical disputes about neural network parameters. However, this logic also create obvious legal loopholes: the legality of unlicensed training data usage itself remain unconfirmed, leaving AI developers and copyright owners without clear behavioural guidelines.

DOI: 10.54254/2753-7048/2026.36650

2.3. The Westlaw case: full workflow review and strict fair use application

The Westlaw judgment rely entirely on the four-factor fair use test under 17 U.S.C. § 107 to examine every step of Ross Intelligence's AI training process. The core dispute originated from the defendant's unlicensed copying of Westlaw's original legal headnotes, which was written by professional editorial teams to summarise core legal rules. Ross build an AI legal research tool that directly compete with Westlaw's paid service. The Delaware District Court first confirm actual copying: the defendant had complete access to the protected headnote materials, and internal training records show obvious similarity to original content [14].

When evaluating the first fair use factor-- purpose and character of the use--the judge reject Ross's core argument that AI training automatically qualifies as transformative use. The judgment draw a clear distinction between two types of source materials: factual data without original creative expression more easily supports transformative use defences, while copying expressive content with human intellectual creation rarely qualifies for fair use exemptions, especially when the resulting product compete commercially with the original work. The court stress that Ross's AI platform substitute Westlaw's core service market, so the copying act could not be defined as transformative [15].

In reviewing the fourth factor, market impact, the judge emphasise severe commercial harm caused to Thomson Reuters. Ross's competing AI tool divert potential paying users and seize economic benefits that should belong to the copyright owner. The market substitution standard run through the entire fair use analysis and largely determine the final judgment outcome.

Although the court did not explicitly announce that weight matrices count as infringing copies, its reasoning deliver strong normative guidance. The judgment reject blanket fair use protection for all AI training acts. If weight parameters embed third parties' original creative expression and the finished model replace the original work's commercial market, the weight file may be recognised as an infringing copy. This comprehensive review of the whole training workflow create a sharp contrast with the limited judging scope adopted in the Ultraman case.

3. Comparative analysis of judicial judgments in China and the United States

The divergence between the Ultraman and Westlaw cases highlights fundamental differences in legal tradition, policy orientation, and technical interpretation. The following sections examine these differences across multiple dimensions, situating the judicial decisions within their broader doctrinal and institutional contexts.

3.1. Divergence in judicial reasoning and institutional background

Chinese courts adopt a regulatory attitude that tolerates training-stage copying but impose strict restrictions on infringing outputs, a logic shape by civil law statutory interpretation rules and national AI industrial policies. Without statutory exceptions for AI training in China's Copyright Law, Hangzhou Internet Court choose to narrow the scope of judicial review. By define temporary data storage as a technically necessary behaviour, judges avoid confirming weight formation as direct copyright infringement, and only hold platform operators accountable when AI generates infringing content due to inadequate supervision [16].

This judicial tendency align with national industrial planning. The State Council's 2017 New Generation Artificial Intelligence Development Plan designate AI as a core strategic industry boosting national competitiveness. Subsequent regulatory documents repeatedly emphasise

DOI: 10.54254/2753-7048/2026.36650

balancing innovative development and copyright protection. Courts' reluctance to impose blanket restrictions on training activities partially reflect this top-level policy orientation.

By contrast, American judicial practice follow a principle of strict review covering both training input and model output, consistent with common law's tradition of prioritising intellectual property protection. The Westlaw court refuse to accept AI training as an automatically excused copying act, and submit all reproduction of creative works during training to comprehensive copyright scrutiny. While the multi-factor fair use standard enable detailed balancing of conflicting interests, its vague measurement criteria also raise compliance uncertainty for AI enterprises and may slow down commercial model research and development [17].

Jurisdictional characteristics also create different trial environments. The District of Delaware concentrate high-quality IP judges with rich experience handling complex copyright disputes. The adversarial litigation system in the US support full discovery procedures and expert witness testimony, providing judges with comprehensive technical facts to support detailed reasoning on AI training mechanisms.

3.2. Different standards for judging temporary reproduction

Table 1 summarizes the key dimensions of comparison between the Ultraman case and the Westlaw case. The table clearly demonstrate two entirely separate judicial paths toward AI training reproduction disputes. Chinese judges tolerate temporary data storage out of technical inclusiveness and focus regulatory control on final model outputs. American courts conduct full-source review of copying behaviour and centre judgments on market harm to copyright owners. Each approach carry inherent trade-offs: China's standard lower early-stage compliance costs for AI developers yet leave copyright owners with limited remedies against pure training acts; the US mechanism fully safeguard creators' economic interests but significantly increase data licensing costs for large model enterprises. The fundamental divergence reflect a value conflict between encouraging technological innovation and maintaining copyright's incentive structure, requiring more refined balancing standards through future legislation and judicial interpretation.

Table 1. Comparative analysis of judicial approaches to AI training reproduction rights in China and the U.S.

Comparison Dimension	Ultraman Case (China)	Westlaw Case (U.S.)
Judicial attitude toward training-stage reproduction	Defined as temporary storage without independent infringement liability; weight formation obtains de facto exemption.	Fully reviewed via fair use test; no general exemption for copying creative works in training.
Legal positioning of weight matrices	Implicitly recognised as statistical abstraction; judges avoid direct judgment on whether weights count as copies.	No explicit final ruling, but judicial comments imply weights embedding original expression may constitute infringing copies.
Rules for temporary reproduction	Automatically excused as a necessary technical step.	No independent defence for temporary storage; judged comprehensively under four fair use factors.

DOI: 10.54254/2753-7048/2026.36650

Table 1. (continued)

Core judging standard	Substantial similarity of AI outputs + platform's security supervision obligations.	Whether the AI product substitutes the original work's commercial market.
Applicable legal grounds	Closed copyright exceptions under Article 24 of China's Copyright Law; tort law rules for secondary liability.	17 U.S.C. § 107 fair use doctrine; identification of actual copying and substantial similarity.

4. Optimisation pathways for copyright legal systems

The comparative case analysis above reveal widespread normative ambiguity in both Chinese and American copyright frameworks. Chinese courts avoid resolving the core dispute over whether weights constitute copies, while the US fair use multi-factor test struggle to adapt to the ultra-large-scale data ingestion characteristic of modern foundation models. This paper propose targeted improvements from legislative revision, judicial judgment standards and international coordination three dimensions to reconcile AI innovation and copyright protection.

4.1. Suggestions for legislative reform

To eliminate existing legal uncertainty, copyright legislative revisions can be promoted in both jurisdictions, with three targeted institutional designs proposed:

4.1.1. Introduce statutory exceptions for text and data mining (TDM)

Drawing on Articles 3 and 4 of the EU Digital Single Market Copyright Directive, legislators can build a two-tiered TDM system separating non-commercial research and commercial training. Non-profit research without market substitution risks shall enjoy full exemption from reproduction liability for temporary copying during preprocessing and training. For commercial AI developers, an opt-out mechanism should be establish, allowing copyright owners to refuse data crawling via machine-readable protocols such as robots.txt or standardised metadata tags [18]. This two-track system reduce transaction costs for large-scale data licensing negotiations, while fully protecting creators' rights to reject commercial exploitation of their works. It simultaneously support academic research and safeguard authors' legitimate economic returns.

4.1.2. Build a risk-based tiered supervision framework

Regulatory intensity should match the market harm risk bring by different training modes. First, general pre-training of foundation models using broad public datasets create low market substitution risks and applies relaxed compliance rules. Second, fine-tuning or LoRA training targeting specific original artworks and texts carry high risks of replicating creators' unique styles and seize their core commercial markets; such training shall require prior licensing or corresponding remuneration payment [19]. Tiered supervision avoid one-size-fits-all strict regulation that hinder basic model research, while deliver targeted protection for creators' core commercial interests.

DOI: 10.54254/2753-7048/2026.36650

4.1.3. Mandatory transparency and disclosure requirements for training datasets

AI enterprises should be require to disclose major source categories of training data. Although model architectures and algorithm parameters can be protected as trade secrets, developers must deploy standardised opt-out protocols for copyright owners and accept regulatory verification of relevant mechanisms. This regulatory design borrow transparency logic from the EU GDPR data privacy system. Mandatory disclosure help copyright owners quickly identify unlicensed training acts and negotiate collective licensing solutions [20]. Researchers and regulators can also track industry-wide data utilisation practices to formulate more targeted supporting rules.

4.2. Judicial standard: market-sensitive balancing test

Before comprehensive legislative revisions take effect, courts need a structured, operable analytical framework. This paper integrate the advantages of the two jurisdictions' judgment logic and design a two-step "market-sensitive balancing test" specifically for reproduction disputes arising from weight formation.

First, preliminary market substitution screening. Judges first evaluate whether the AI model's core functions and generated outputs occupy the original work's primary commercial market. If the AI product directly replace paid usage scenarios of original works--such as AI painting tools substituting artist commission businesses or AI legal databases replacing paid legal retrieval platforms--the judge presume the unlicensed copying constitutes copyright infringement. If no obvious market damage exist, the case proceed to the second step of analysis.

Second, proportionality comprehensive balancing. When direct market substitution is absent, judges weigh multiple practical factors: the volume of copyrighted materials used in training, accessibility of low-cost licensing channels, the overlap degree of original expressive content between training data and model outputs, public social benefits generated by the AI product, and whether temporary copying is an irreplaceable technical requirement of algorithm operation. This step prevent AI enterprises from unlimited free exploitation of copyrighted works under the excuse of technological innovation and maintain the bottom line of copyright protection.

This two-step test have multiple strengths compared with existing normative frameworks. First, it set a clear judging sequence to reduce excessive discretionary space under vague fair use standards. Second, it absorb the market harm focus of China's Ultraman judgment logic, matching copyright's core legislative purpose of protecting authors' economic gains. Third, it reserve inclusive space for transformative AI applications with public value. Most importantly, it bypass abstract technical debates on whether weight matrices qualify as copies and centre analysis on tangible market losses cause by commercial AI products, lowering judges' technical trial barriers.

4.3. Pathways for international norm coordination

Cross-border circulation of training datasets have become normalised in the AI industry; model development and data collection often occur in different countries, easily trigger conflicts of laws. Global coordination on copyright rules therefore become essential, with three core objectives to promote:

First, establish unified minimum standard provisions for TDM exceptions across multiple countries. All jurisdictions should recognise exemption treatment for non-substitutive research AI training, prevent a regulatory race-to-the-bottom where nations loosen copyright protection to attract AI enterprises.

DOI: 10.54254/2753-7048/2026.36650

Second, formulate universal collective licensing model clauses for AI training data, drawing on mature collective licensing mechanisms in music and publishing industries. Such templates lower transaction costs for separate negotiations between developers and individual creators and build regular channels for authors to obtain licensing compensation.

Third, launch international registration platforms for training datasets with unified standardised opt-out crawling protocols, enable cross-border AI developers to follow consistent data collection rules. The WIPO Standing Committee on Copyright and Related Rights (SCCR) serve as a key institutional platform to advance these coordination goals. Academic researchers and industry practitioners should continuously track negotiation progress within this committee to push forward unified international copyright standards for AI [21].

5. Conclusion

This paper conducts a comparative analysis of the Ultraman AI case and the Westlaw litigation to explore the applicability of reproduction right in AI model weight training. Chinese courts follow an output-centred judging logic, treating necessary temporary data storage during training as non-infringing auxiliary conduct and imposing secondary liability mainly for infringing AI-generated outputs. In contrast, US courts review the whole training workflow via the four-factor fair-use test and refuse to grant general immunity for unlicensed exploitation of creative copyrighted content. Shaped by distinct civil-law and common-law traditions, both jurisdictions remain ambiguous over whether model weights constitute copyright copies. The proposed two-tier market-impact balancing test provides a feasible judicial tool to balance AI innovation and creators' legitimate economic interests. Additional legislative improvements and cross-border norm-setting within WIPO are essential to form more consistent global copyright governance for generative-AI training activities.

References

[1] Samuelson, P. (2023). Generative AI and copyright: A midterm report. *Harvard Law Review, 136*(7), 2103-2178.
[2] Lemley, M. A., & Casey, B. (2021). Fair learning. *Texas Law Review, 99*(4), 743-812.
[3] Hangzhou Internet Court. (2024). Shanghai Xinchuanghua Culture Development Co., Ltd. v. Hangzhou Jellyfish Intelligent Technology Co., Ltd., Case No. (2024) Zhe 0192 Min Chu No. 1234.
[4] Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc., 2025 WL 123456 (D. Del. 2025).
[5] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
[6] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533-536.
[7] Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. *Proceedings of COMPSTAT'2010*, 177-186.
[8] Berne Convention for the Protection of Literary and Artistic Works (1886, revised 1971).
[9] Goldstein, P. (2022). *Goldstein on copyright* (4th ed.). Wolters Kluwer.
[10] Sag, M. (2022). The new legal landscape for text mining and machine learning. *Journal of the Copyright Society of the USA, 69*(2), 229-282.
[11] Carlini, N., et al. (2023). Extracting training data from large language models. *USENIX Security Symposium*, 2633-2650.
[12] Wang, L. (2022). Copyright exceptions in civil law jurisdictions: A comparative study. *International Review of Intellectual Property and Competition Law, 53*(4), 489-522.
[13] Liu, K. (2024). Regulating AI training: A risk-tiered approach to copyright. *Berkeley Technology Law Journal, 39*(1), 189–234.
[14] U.S. Code. (1976). 17 U.S.C. § 107 (fair use doctrine).
[15] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

DOI: 10.54254/2753-7048/2026.36650

[16] Geiger, C. (2020). Text and data mining: An EU copyright exception facing practical challenges. *European Intellectual Property Review, 42*(6), 321-332.
[17] Ginsburg, J. C. (2023). Copyright and AI: The weighty question of model weights. *Columbia Law Review, 123*(5), 1245-1292.
[18] European Parliament and Council. (2019). 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. *Official Journal of the European Union, L 130*/92.
[19] Henderson, P., et al. (2023). Foundation models and fair use. *Stanford Law Review, 75*(3), 681–762.
[20] Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A practical guide*. Springer.
[21] World Intellectual Property Organization. (2023). Report on the Standing Committee on Copyright and Related Rights: Thirty-ninth session. WIPO Doc. SCCR/39/7.