Proceedings of the 5th International Conference on International Law and Legal Policy
DOI: 10.54254/2753-7048/2026.37121

Fair Use Boundaries for AI Training Data: A Comparative Study of Chinese and U.S. Copyright Law

Xurong Sun

Language and Culture College, South China Business College, Guangdong University of Foreign Studies, Guangzhou, China
ngjwc@gwng.edu.cn

Abstract. With the advent of generative artificial intelligence, there has been a significant amount of legal limbo surrounding the use of copyrighted material for training purposes. This paper explores the parameters of acceptable AI training as outlined by both Chinese and U.S. copyright law by analyzing recent court rulings. It highlights three situations: Kadrey v. Meta Platforms, Inc. and Bartz v. Anthropic PBC in the United States, and the Ultraman AIGC platform case in China. The method used is doctrinal legal analysis, comparative legal analysis and case study. It claims that the two countries have different institutional approaches to the training of AI: the United States with the flexible fair use doctrine of 17 U. S. C. § 107, China with the enumerated limitations and general conditions of Article 24 of the Copyright Law. However, it can be challenging for both systems to adapt the market impacts of AI to copyrightable standards. Judgements on the competition of AI-generated output or AI services to the market for original copyrighted work(s) are scarce. This paper does not suggest a new legal test, but rather that the test of market substitution should be given more explicit and structured consideration in the existing copyright doctrines. This may help create more consistency in analysis and allow for more flexibility in technology innovation.

Keywords: fair use, artificial intelligence, AI training data, copyright law, comparative law

1. Introduction

As generative AI becomes more prominent, the dynamic between technology and copyright is undergoing transformation. Large language models and other such systems require a lot of data to be processed, including books, articles, images, music and other copyright-protected materials. What this means then is that the fundamental legal issue is not whether an AI system can create new pieces of work. Whether the copying that occurs in training is lawful under existing Copyright Law.

This is now a pressing problem in the United States. In June 2025, the U. S. District Court for the Northern District of California ruled on two important cases involving AI training. The court in Bartz v. Anthropic PBC ruled that copying books for training LLM was fair use; but Anthropic's additional acts of acquiring and permanently holding pirated books in a central library were not [1]. On the record before the court, another judge in the same district ruled the same way two days later:

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.37121

Meta's use of copyrighted books to train its Llama model was fair use [2]. While both decisions were generally similar, they differed in their emphasis on the transformative use vs. market harm.

The situation in China is very different. Under Article 24 of the Copyright Law, some uses may be made without permission or payment, provided that they comply with certain requirements – and most importantly the use must not prejudice the normal exploitation of the work or the legitimate interests of the copyright owner [3]. In recent AIGC litigation, such as the 2024 Ultraman cases, Chinese courts have started to apply such provisions to generative AI. However, unlike the U. S. Copyright Act, there is no broad exception like §107 in China's legislation, and no specific exception for AI training.

This paper therefore aims to provide answers to three related questions. First, what are the boundaries of "permissible" AI training, as defined by U. S. and Chinese courts under their respective copyright systems? Second, why is there a different approach to the problem in the two systems? Third, and finally, do both systems face the same root challenge when it comes to copyright issues in the context of using AI: when does use become problematic for the markets copyright is intended to safeguard?

The central argument is this: do not assume that the two jurisdictions will have the same rules. The U.S. system, which has an open-ended doctrine of "fair use", provides more flexibility. There is greater certainty in China's model, which has more rigid statutory boundaries. However, both approaches lack clarity regarding the link between the training of AI and the outputs it generates, and market replacement. Market substitution should not be an alternative to existing fair use doctrines, or a new stand-alone test. Instead it should be treated more explicitly and structured in the frameworks for the analysis of market harm and legitimate interests.

2. U.S. Copyright Law fair use boundaries

In some respects, the most flexible mechanism for dealing with AI training is the U. S. approach which does not entail an exhaustive list of approved uses. Section 107 provides four factors for courts to consider on a case-by-case basis. That flexibility is an asset: it allows courts to use principles that have been formulated over the years to technologies that Congress never envisioned. But that is also the weakness: the legal significance of the training of the AI is ultimately contingent on the way that courts define its function, the connection that exists between the original works and the AI, and the impact of the AI on markets. The following two sections examine the impact of this in Kadrey and Bartz, two cases that tackled a similar technology challenge in markedly different ways.

2.1. The fair use doctrine

There is a fundamental principle of the Copyright Act, Section 107, that some uses of a copyrighted work may qualify as "fair use" instead of being considered infringement, such as for criticism, comment, news reporting, teaching, scholarship, or research. However, that's not all. The provision also directs courts to consider four factors, including the purpose and character of the use, the nature of the copyrighted work, the quantity of the material used and most significantly, the effect of the use upon the potential market for the copyrighted work [4].

This first factor has become a determining factor in the answers to the other three factors, in no small part due to the transformative-use doctrine. In Campbell v. Acuff-Rose Music, Inc. however, the Supreme Court affirmed that a use is more likely to be fair when it "creates something new or

DOI: 10.54254/2753-7048/2026.37121

different" from the original work, and thus is not a replacement for it [5]. That is, how the copying is being used is more important than whether it was actually copied.

However, the issue of AI training introduces a lot of complications. During a typical training session, the model could input the complete text of thousands of books, and regenerate them word-for-word. But, those copies are not being served to the readers, they are not replacements for the originals. Rather, the text is used as raw data that the model uses to detect patterns in language and enhance performance. Therefore, here is the hard question: Is that sort of copying to be considered something different from expressive reproduction, namely technical analysis? Is it still considered to be communicating or leveraging value from the original works? The answer will be significant because the description of the copying will determine the understanding of the purpose and the character of the copying under factor one. The fourth factor complicates things. Section 107 makes it quite clear that courts must take into account the impact that the defendant's use has on existing and potential markets for the work. Now is where things start getting really difficult with AI. Where a model learns from copyrighted books, and begins to produce outputs that are now available in the market that compete with those books, the market harm is no longer hypothetical, but real. The U.S. Copyright Office's 2025 report states as much: AI training can "materially harm" the market by (1) AI-produced works supplanting training materials or (2) by undermining existing and nascent markets for similar creative works [1]. Thus, even when the purpose of training has a tendency towards the fair use end of the spectrum, the market effect cannot be ignored as a part of the statutory analysis.

2.2. Case study: Kadrey v. Meta Platforms

The Northern District of California recently ruled in favor of Meta on the fair-use issue and entered summary judgment on that question, which was whether it was permissible to use copyrighted books in training LLaMA models on June 25, 2025 [2]. Meta's use was considered by the court to be very transformational. After all these books were not being reprinted or acted out to readers. They were being used as training inputs, or raw material, to help the model learn to produce language.

However, the importance of Kadrey is not just that the court recognized AI training as transformative. It is also very interesting on how it deals with market harm. In its record before it, plaintiffs were not able to demonstrate that Meta's use of their books was actually displacing demand. That distinction puts an important qualification on the transformative-use analysis. While a technological or analytical use may weigh in favor of a finding of "fair use", that does not mean that the economic consequences are not important. The fourth factor may have a great deal more significance if copyright owners can demonstrate that the outputs of an AI system are now indeed in competition with the markets for their original works.

However, in this respect Kadrey leaves the door open. While the court recognised that the harm to the market may be significant, it did not provide a general framework for determining when AI-generated outputs have replaced the need for the use of copyrighted works to a legally relevant standard. That is a very big issue in the AI world. Market displacement is difficult to demonstrate during the training phase, prior to widespread deployment of the system and before good consumer behavior data exists. So Kadrey provides us with a valuable initial point for discussion regarding the harm to the market, but it does not yet provide us with a full-blown approach for assessing market substitution in AI related fair-use cases.

DOI: 10.54254/2753-7048/2026.37121

2.3. Case study: Bartz v. Anthropic

The same court heard a related dispute two days prior on June 23, 2025 in Bartz v. Anthropic PBC. In his decision, Judge William Alsup found that Anthropic's use of the copyrighted books to train its language models was fair use because the training process was "exceedingly transformative" [3].

However, there is something about Bartz that makes the holding really interesting. The court took a careful approach in defining the difference between using copyrighted works for training purposes and how the works were acquired and retained. Some of the books were legally obtained by Anthropic. It had also downloaded millions of pirated copies, and kept them in a permanent, central library. While the court ruled the training was fair use, the same could not be extended to those unauthorized library copies [3]. That distinction matters. It establishes that the transformative inquiry cannot be addressed in the theoretical vacuum as to whether AI training is transformative. The context also needs to be considered - the process of obtaining, storing and using the materials.

The treatment of market competition also shows a slight but definite difference between Bartz and Kadrey. The court in Bartz did not appear to be as keen to rule that AIs could create an output that is potentially a competitor to a human-made work on its own. While Kadrey, by contrast, emphasized the importance of "dilutive market harm" and urged that any fears of this type of market harm be based on evidence rather than on speculation [2, 3]. However, these two cases do not necessarily contradict each other. They do contain varying inclinations regarding the balance of AI outputs and copyrighted products markets.

Overall, Kadrey and Bartz illustrate the robustness and insecurities of the U. S. fair-use doctrine. The case-by-case approach allows for a flexible interpretation to recognize that there are various forms of copying based on AI, to which different purposes, acquisition methods, and market effects could apply. However, methodological problems remain with this flexibility: What is the evidence to show that the use of AI has gone beyond merely creating a new function of a technological tool to replacing the market for the original works? In the following section, on the Chinese approach, that question proves all the more relevant.

2.4. Assessment of the U.S. judicial way of AI instruction

It has much to recommend it, the U. S. approach. New technologies not considered by Congress in 1976 can be accommodated by Section 107. Protective doctrine of "transformative use" provides judiciary with a measure for distinguishing copying uses from uses that are truly different. The fourth factor--market effect--offers a statutory opportunity for considering economic implications [4].

This flexibility comes at a cost, however. It has uncertainty attached and that's a compromise. While both Kadrey and Bartz considered the use of AI training to be fair use, they did not view market competition and data acquisition in the same light [2, 3]. What does this reflect? Howevver, the fate of AI fair-use doctrine is not going to depend on whether the use is simply "transformative" --courts will also have to determine: what markets matter? What do you mean by evidence of market harm? But when does the competition of AI-generated content get to the point of replacing it?

That is no simple set of questions, particularly due to the fact that AI breaks down the traditional lines. A book being sold to readers is not being sold in the same form as it is being used for training. However, under the right circumstances, an AI output may be able to do something similar to compete for consumer demand. The U.S. framework provides some guidance for courts to consider these effects, although the recent cases have not yet provided a consistent approach as to when these effects become relevant.

DOI: 10.54254/2753-7048/2026.37121

3. Fair use boundaries under Chinese Copyright Law

The Chinese framework differs from the U.S. fair-use system in two respects: first, the statutory framework, and second, the role of judicial discretion. The exceptions contained within Article 24 of the Copyright Law outline specific circumstances where a copyrighted work can be used without permission, but are also subject to "general conditions", "the use must not interfere with the normal exploitation of the work" and "the use must not unreasonably prejudice the legitimate interests of the copyright owner" [5]. The Chinese system of balancing, unlike U. S. fair use, is not a general statutory provision, but rather an institutional mechanism that is more specific.

That gap has an impact--particularly for AI training. Large scale training frequently entails the reproduction, processing and analysis of copyrighted materials in ways that do not fit squarely into the old-fashioned categories listed in Article 24. So it's not only if AI training is technologically new. Whether or not the current statutory scheme would provide courts with sufficient flexibility to fit the new use in. The Ultraman lawsuit is a good example of this tension. However, instead of establishing a new statutory exception for AI training, Chinese courts discussed the conduct of AI with respect to copyright and the platform's obligation. So the case is a good illustration of the flexibility and limitations of the existing framework.

3.1. Article 24 and the structured copyright limitation framework

The "without permission" exception in Article 24 of the Copyright Law of the People's Republic of China (PRC) provides examples of those exceptions, and imposes general limitations on normal use and the legitimate interests of the copyright holder [6]. More determinate than the U.S. fair-use system, but AI training is not provided as a standalone exception.

That is a problem. AI training might involve mass reproducing and processing copyrighted content, but the purpose for that copying might not be as it is in other traditional uses such as quotation, teaching, or personal use. Thus, whether training can be accommodated within Article 24 will be very much dependent on how courts define the act and the purpose of the act.

The institutional variation between the two systems is of no less importance. The U.S. courts have been instructed in §107 to consider a number of factors in determining what is a "fair" use. The Chinese courts, on the other hand, work in specified statutory situations and thus have limited scope of creating a broad, common law-like fair-use doctrine. But that does not rule out the possibility for the Chinese judicial organs to react to technological change; adaptation, however, should remain closer to the statute's categories and restrictions provided in the Copyright Law.

3.2. China's AI governance framework and its interaction with Copyright Law

Copyright Law is not the only response that China has had to generative AI. The Interim Measures to Regulate Generative Artificial Intelligence Services, which came into force in August 2023, laid down provisions for content management, security, and generative-AI service providers' responsibilities [7]. While, these measures are applicable to the general governance of generative AI, they should not be interpreted as a general exception to copyright for the use of copyrighted works as training material. This distinction is important because AI copyright issues can occur at various points in the development. Relevant questions at the input/training stage are the acquisition, reproduction, processing and analysis of copyrighted materials. The attention is now at the output stage for generation, dissemination and potential reproduction of protected expression. While these are related, they are not necessarily the same.

DOI: 10.54254/2753-7048/2026.37121

By maintaining the analytical separation of these stages, one also prevents a conclusion that is too general, such as that everything that comes out is lawful, because it was trained lawfully. On the other hand, a violative response does not necessarily imply that all actions taken throughout the training were necessarily infringing. Careful copyright analysis should determine the specific act or acts in question, as well as the legal interest that the act or acts infringes.

This distinction is relevant, as AI copyright conflicts are at least two-fold: Input and training (the acquisition, reproduction, processing, and analysis of copyrighted materials); and Output and dissemination (the generation, transmission, and public availability of potentially infringing materials).

This approach gives a court the flexibility to take into account that not all training is necessarily unlawful, and that if an output is later found to be unlawful, this does not mean that all uses during the training process were unlawful as well.

3.3. Case study: the Ultraman AIGC platform case

An important Chinese case to consider these questions is (2024) Zhe 0192 Min Chu No.1587, which was affirmed on appeal as (2024) Zhe 01 Min Zhong No. 10332 (hereafter referred to as "the China case") [8]. The controversy is about an AI platform that offers image-producing and model-related services. The dispute ultimately hinged on the platform's responsibility for the generation and dissemination of allegedly infringing content as copyrighted Ultraman-related materials were used in the operation of this platform [9].

The focus of the courts' analysis was primarily on the responsibilities of the platform, and whether the platform had taken reasonable steps in response to materials that were potentially infringing. But it remains relevant for AI training as the judicial reasoning makes a distinction (at least analytically) between the processing of materials as part of the AI process, and the creation of potentially infringing outputs.

The distinction should be expressed with caution. The Ultraman case did not establish a general exception to allow AI systems to be trained on copyrighted works. Nor does it define something like AI training as reasonable use in and of itself just because the nature of the use is in connection with analysis and not expression. Instead, the case provides an example of a new judicial nudging towards reviewing AI-related behavior in the light of the specific purpose of the action and the impact on copyright holders [10]. It is important for the current comparison. At this moment, it remains unclear how AI training can be realized in a specific statutory way in the Chinese framework. However, some analytical tools exist for courts to differentiate various stages and functions of AI-related conduct in the shape of existing copyright principles. What remains to be seen is the degree to which such judicial interpretation can go without the effect of de facto creating a new copyright exception through case law.

3.4. Evaluation of the Chinese judicial approach to AI training

The Chinese system provides a more structured statutory framework than U. S. fair use, which also has a down side: less flexibility for the courts. Article 24 provides identifiable statutory boundaries for courts and the Ultraman litigation illustrates how these boundaries can be interpreted in light of technological developments. Meanwhile, the lack of an express exception that expressly covers

AI training process involves a lot of uncertainty for copyright owners and AI developers. It is not just that there is no text and data-mining exception. At a more basic level, Chinese law has yet to establish a uniform approach to the economic impact of AI-related uses. Yet the underlying ideas--

DOI: 10.54254/2753-7048/2026.37121

namely, "normal exploitation" of a work and protection of the copyright owner's "legitimate interests"--might well offer a framework for addressing market harm concerns. While, current case law does not provide a sufficiently systematic approach to deciding whether the content generated by AI is competing with or supplanting the market of the works.

That gives a good comparison to what it is like in the United States. This is a problem that the two systems take different doctrinal paths. China does not tackle the matter of market effects directly in the law, but rather through the concept of statutory limitation of normal exploitation and legitimate interests. U.S. law explicitly reference market effects through §107. Yet, both of these frameworks have not yet given a clear indication of when AI activity would lead to legally relevant market substitution.

4. Comparative analysis: alternative methods and common issues

The analysis to date has outlined some significant discrepancies in the U. S. and Chinese approach to AI copyright conflicts--both in terms of the statutory framework and the reasoning of the courts. However, such variations do not render comparison moot. Sure, the selected cases are from various factual settings. As for the institutional question, however, they all struggle with something of a common issue: what should Copyright Law do when copyright-protected works move into technological systems capable of producing copyright-protected works on an order of magnitude that Copyright Law did not account for?

That is, the comparison is not to try to make the cases the same precedent. This is not to say the same legal test should be used in both locations. Instead, the analysis focuses on the way the two systems describe the uses that involve AI, whether and how they consider the copyright and technological interests, and whether and how they consider the economic consequences for copyright owners. From this perspective, the cases show a substantial doctrinal split as well as a common, albeit limited, challenge: they have yet to devise a methodology that is stable in the translation of AI-related market effects into copyright analysis.

4.1. The comparability of the selected cases

The selected cases are similar but at a certain level of analysis. They are not to be regarded as being somehow on the same level as factual disagreements. Kadrey and Bartz both directly discuss copyrighted books used for training large language models, so they provide the most direct insight into the way that U.S. courts view the use of copyrighted materials in the context of the fair-use doctrine [2, 3]. By contrast, the Ultraman litigation is based on a different set of facts. The focus is on the responsibility of an AI platform for the infringing output and the difference between the AI's generation and dissemination and the training process. Therefore, it can not be considered as a Chinese version of Kadrey or Bartz.

The cases do offer a viable ground for comparison--for three interrelated reasons. All of them are about the relationship between copyrighted works and AI related technologies. They all involve courts enforcing copyright concepts upon technologies that were not expressly envisioned in the rules of the Copyright Law. All of which point to the wider policy conundrum that is the protection of the economic interests of copyright holders against the further development of technology. So it's more of a functional comparison to fact.

Kadrey and Bartz demonstrate an open-ended fair use doctrine's direct response to AI training. The way the statutory system tackles AI generated conduct, relying on the current copyright framework in a more structured approach is shown in Ultraman. Comparing them is not about what

DOI: 10.54254/2753-7048/2026.37121

is "right" or "wrong" about either legal system, but about what each system makes visible and what each system leaves under-developed.

4.2. The divergence between the two approaches

The fundamental difference is in legal structure. The U.S. framework is open-ended: Section 107 lets courts weigh four factors, and concepts like transformative use can evolve through precedent [4]. That gives courts a lot of flexibility in responding to technological change. China's framework is more structured: Article 24 identifies specific permissible circumstances and imposes general constraints around normal exploitation and legitimate interests. The result is greater statutory predictability, but less room for courts to create new exceptions through individual cases.

The difference can be summarised in Table 1. The divergence therefore does not simply reflect different attitudes toward AI. It reflects the institutional design of the two copyright systems.

Table 1. Comparison of the United States and Chinese copyright frameworks for AI-related use

Dimension	United States	China
Core framework	Fair use under §107	Article 24 limitations
Judicial flexibility	High	More constrained
Transformative use	Developed judicially	Not an equivalent independent doctrine
AI training	Can potentially qualify as fair use	No express AI-training exception
Market effects	Expressly relevant under factor four	Relevant through normal exploitation and legitimate interests
Main uncertainty	Evidence and measurement of market harm	Statutory basis and scope of permissible training

4.3. Market substitution as a shared analytical challenge

The bottom line from the comparison--and it is not that both systems simply ignore market substitution. It would be an overstatement. Market harm was explicitly a factor for Kadrey and Bartz [2, 3] was a bit more cautious about talk of competitive displacement. In fact, Chinese law on copyright already focuses on normal exploitation and legitimate interests as set forth in Article 24 of the Copyright Laws [8]. However, what is the real issue? It is more slender than a full blind spot. A systematic, workable approach for translating market effects of AI into copyright analysis is still lacking, both in either jurisdiction. Market factors are not ignored, but are not linked with well-defined rules on when they are relevant for legal purposes.

Consider how copyright infringement cases are normally handled. They typically are related to fairly easy-to-understand markets. A pirated book will directly compete with the original book. An unauthorized photo decreases the demand for a licensed photo. Pretty straightforward. AI clashes are dirtier. A model could be trained on millions of copyrighted works, and it won't duplicate a specific one. However, the same model could produce outputs that are in the same kind of commercial space later. Thus, the economic impact can manifest itself at various levels.

DOI: 10.54254/2753-7048/2026.37121

For analysis purposes, it is possible to identify three forms of potential market effect. In direct substitution, the AI produced output is so close to a specific copyrighted work that consumers opt for the output instead of the original. This is a case with the greatest claim of market harm, the linkage between the challenged use and the original work is relatively direct. Market dilution is a more general competitive phenomenon: AI systems create a lot of content competing with a more general class of creations in the training set.

This is more difficult to determine because the connection between any specific copyrighted work and the claimed economic harm is less obvious. For copyrighted materials, there are concerns about the potential for widespread unlicensed use for training AI systems, which could impact the value of a potential licensing market. Extra care needs to be taken with this one. Under the theory of copyright, if there is a potential market in the future for a right-holder to license the use of a work in a new technology, copyright might effectively put the right-holder in a position of controlling the use of it.

They are not legal categories, created by the courts. They're a framework for analysing various types of market impact. The more similar the allegedly caused harm is to the original harm the more likely that the connection to the evidence will be strong. Claims of market harm more distant from the market require more care, particularly if the alleged market has not yet reached the level of an established form of exploitation. Therefore, the central issue is not whether there has been "market substitution". Whether the alleged substitution is direct or indirect; whether it is a significant share of the market for the protected interest or just a minor copy; whether it is an interest that Copyright Law is properly intended to protect. This makes sure that the market effects remain within the existing copyright assessment and that market substitution is not an autonomous test.

4.4. Proposals for a future AI copyright governance

There is no evidence that the comparative analysis would justify a new market-substitution test to replace fair use and the current copyright limitations in China. Rather, there are three smaller reforms to be given preference.

4.4.1. Judicial standards for market substitution in the U.S.

Existing four-factor test should be retained by U. S. courts, and should be given more structure when assessing market harm caused by AI.

Courts should make a distinction between: direct substitution for specific copyrighted works; more broadly dilution of markets for similar works; and hypothetical licensing markets. The farther away from the original work the stronger the evidence should be.

Furthermore, this would maintain the flexibility of §107 and remove some of the ambiguity about what factors constitute "dilutive harm" under the current law, as the Copyright Office has already noted in its 2025 report that AI-generated content can be a direct substitute for training works, or it can dilute the market for similar works [1].

4.4.2. China's legislative expansion or TDM exceptions

China has another institutional issue. More transparency is needed on the market effects of AI training to help courts formulate a consistent approach to assessing whether and when AI training is permissible under current Article 24 restrictions.

DOI: 10.54254/2753-7048/2026.37121

This could be a specific text and data mining exception or another narrowly framed statutory provision to deal with AI training. This could be based on various factors such as the intended use of the training, the legal acquisition of the underlying works, commercial vs. non-commercial uses, protective measures for AI-generated content, and clearly demonstrated harm to copyright holders.

Legislative action is institutionalized in a way that is an advantage over relying solely upon individual judicial decisions. Since Article 24 is framed in terms of exceptions, there is a huge possibility of case law to expand the scope of the exceptions, which might result in creating uncertainties as to the proper parameters of judicial interpretation. Legislative clarification could also be in the form of setting lines of demarcation and letting courts draw the lines between them in specific technology cases.

4.4.3. Comparative perspective: moving towards more convergence

It is important to note that full harmonization between the two systems is impractical, and unnecessary, for the U. S. and Chinese systems. It is likely that the United States will continue to have a more flexible "fair-use" interpretation, whereas China will continue to have a more structured statutory interpretation. Comparative convergence is more realistically sought for at the level of questions than identical legal rules.

Far more realistic is convergence on the questions that courts should ask: (1) What is the use for which the AI is used? (2) How transformative or analytical is that use? (3) What was the process for acquiring copyrighted materials? (4) Is there reproduction/exposition of protected expression by the AI system? (5) Can the created AI product replace the original works? (6) Is the so-called market a genuine market and copyright protected?

This convergence of procedure would make it more predictable without having to introduce the same substantive rules in the two systems.

5. Conclusion

With large-scale training, generative AI has thrown up some genuine ambiguities about the scope of traditional copyright concepts, which both sides of the Pacific are currently tackling. This paper has examined recent case law, such as Kadrey v. Meta Platforms, Bartz v. Anthropic, and the Hangzhou Ultraman case, in the U.S. and China.

One thing is clear from the comparison: The two jurisdictions deal with copyright issues relating to AI in a very different legal framework. The U. S. courts operate under the "flexible" standard of fair use", 17 U.S.C. §107, which includes a consideration of purpose and character, in which transformative use is a major factor, as well as an express reference to market harm in the fourth factor. However, China operates under the more structured provision of Article 24 of the Copyright Law, which lists the uses that are permissible, and imposes restrictions on normal use and on the legitimate interests of the copyright holders.

However, all those differences aside, the cases present a more limited common challenge. Until now, there is no systematic and predictable method, either in either jurisdiction, to find out when AI-related uses generate market effects of legal significance. But, courts have not overlooked the effects of market harm; they haven't. However, the decided cases reveal a lack of clarity on how any such harm is to be identified, assessed and balanced, particularly if the AI generated work is competing with other original works, classes of original works, or potential licensing opportunities.

Thus the correct answer is not to do away with the existing copyright doctrines and substitute them with an independent market-substitution test. A better way to approach this is to provide a

DOI: 10.54254/2753-7048/2026.37121

more structuring analysis of market effects in the existing frameworks, making a distinction between markets for direct substitution, market for substitution in general and speculative licensing markets. That sort of approach would maintain flexibility for the technological innovation that copyright requires, and it would also enhance the power of copyright to support the legitimate economic motivations that drive creative work.

References

[1] U.S. Copyright Office. (2025). *Copyright and artificial intelligence*. Washington Office.
[2] *Kadrey v. Meta Platforms, Inc.* (2025). No. 3:23-cv-03417-VC.
[3] *Bartz v. Anthropic PBC.* (2025). No. 3:24-cv-05417-WHA.
[4] Copyright Act of 1976, 17 U.S.C. §107.
[5] *Campbell v. Acuff-Rose Music, Inc.* (1994). No. 510 U.S. 569.
[6] Copyright law of the People's Republic of China (2020 revision), art. 24.
[7] Qi, L. (2025). Legal risks and institutional supply of AI training data. *People's Judicature, 11*, 11-19.
[8] Shanghai Xinchuanghua Culture Development Co Ltd v Hangzhou Jellyfish Intelligence Technology Co Ltd. (2024). No. Zhe 0192 Min Chu No 1587.
[9] Sun, J. M. (2025). From fair use to fair learning: Copyright dilemmas and normative reconstruction of AI training data. *China Law Review, 5*, 52.
[10] Cyberspace Administration of China. (2023). Interim measures for the management of generative artificial intelligence services. Retrieved from https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm.