

Research on Copyright Governance and Security Challenges of Training Data for Generative Artificial Intelligence

A Comparative Analysis Based on Typical Litigation Cases

Jian Huiyao

*Guangzhou Nanfang College, Department of Big Data Management and Application,
Guangzhou, China*

Email: 3057884729@qq.com

ABSTRACT:

Generative AI depends on training with very large volumes of data, and the acquisition and use of that data has become a focal point for copyright infringement and for privacy and security risks. This paper examines how generative AI training data is governed and what security problems it raises. It compares the reasoning of Chinese and American courts in copyright disputes over training data, traces the concentration of infringement risk at the data-input stage, and considers why the fair use doctrine fits these cases poorly and why licensing and compensation mechanisms no longer work. From there it turns to security, mapping risks across the data life cycle and showing, with reference to membership inference attacks and privacy extraction in federated learning, how model memorization allows training data to leak. Technical responses such as verifiable privacy-preserving inference are assessed alongside these risks. The paper concludes by arguing for a governance framework that pairs procedural proof with substantive rules, adapts fair use to local conditions, uses standardization as a lever, and relies on agile multi-party co-governance. Its purpose is to clarify how innovation and the protection of data interests can be balanced.

Keywords: generative artificial intelligence; training data; copyright governance; fair use; membership inference attack; data security

1. INTRODUCTION

1.1 Research Background

Generative AI is reshaping the technological landscape and the organization of production at remarkable speed. Industry forecasts place the market for AIGC in China at 720.2 billion RMB by 2028; ChatGPT passed one million users five days after release and reached tens of millions within three months [4]. Such figures show how quickly the technology has diffused. Its capabilities rest, however, on the absorption of very large volumes of training data, and it is the acquisition and use of that data that has become the centre of legal controversy. By 31 December 2024, 302 generative AI services had been registered with the Chinese cyberspace authorities, and courts across the country heard 1,659 cases involving AI between 2021 and 2024 [1]. Copyright infringement and privacy leakage are now prominent risks, and reconciling the encouragement of innovation with the protection of data interests has become a central problem for data governance.

1.2 Research Significance

Two questions give the subject its theoretical interest. The first is whether machine learning can be treated as equivalent to human learning; the second is whether non-expressive use can support an exemption. Both require the fair use doctrine to be reconsidered in the light of current practice. The practical interest is no smaller. Security risks attend every stage of the data life cycle, from collection and storage through computation, output, and deletion [6], and membership inference attacks can disclose whether a particular item formed part of a training set [7], exposing personal privacy and trade secrets in new ways. A comparison of leading Chinese and American cases speaks to both concerns.

1.3 Research Questions

Three questions guide the analysis. First, how far do Chinese and American courts differ in their approach to copyright disputes over training data? Second, what problems of copyright compliance and security does training data present? Third, what governance arrangements would protect rights holders without discouraging innovation?

2. LITERATURE REVIEW

2.1 The Substantive Law of Copyright Governance

Fair use sits at the centre of disputes over generative AI training data. Work based on the U.S. Copyright Office report Copyright and Artificial Intelligence treats training as a process of statistical optimization: works must be scanned in full and copied at the level of pixels before statistical patterns can be extracted, a process quite unlike the non-linear cognition of human learning, so that an appeal to "learning" cannot simply be carried over. The same work proposes a market dilution test for the substitution damage that AI inflicts on the potential market for a work, and argues that retrieval-augmented generation (RAG) should not benefit from the fair use exception, since it adds nothing transformative and substitutes for the original [1]. Other writers divide training into three stages, namely data input, model learning, and model optimization, and locate the infringement risk mainly in the first [5]. Their criticism of the non-expressive use doctrine is that it licenses developers to use works without payment, a one-way transfer that leaves creators bearing the costs while developers take the returns [5]. Comparative work reaches a similar impasse: the American transformative use test is open-ended but subjective, the Japanese flexible fair use provisions are confined to information and communication, and the European TDM exception allows rights holders to opt out, which leaves AI firms unsure whether they may use the data at all. Each model has advantages, and none transfers easily [5].

2.2 Security Risks and Privacy Attacks

Applying grounded theory to 38 policy texts, one study derived 212 initial concepts, 54 sub-categories, and 15 main categories, identified risks at the six stages of collection, storage, annotation, computation, output, and deletion, and proposed a three-part governance system of rules, procedures, and technology [6]. Work on membership inference attacks reaches a complementary conclusion. Over-parameterized networks tend to memorize sensitive items in the training set unintentionally, and because a model responds differently to data it has seen, an attacker who can query the model can infer whether a given item was part of its training data [7]. In federated large language models the FedLLM-Legal benchmark and the FedLLM-Extract attack show a semi-honest client extracting personal identifiers from other clients, with coverage

reaching 44.2 percent [3]. Defences have been proposed as well. Vp3CNN, built on VOLE-ZKP and secret sharing, combines privacy protection with verifiable inference [11], and GAN-SOFT and the Entro-DWE ensemble reduce the success of attacks substantially. The emphasis of defence has moved from preventing information from being read to removing memorization from the model itself [7].

2.3 Governance Systems and the Role of Standards

Standardization has been described as the first move in managing the risks of generative AI, on the grounds that it makes technology more controllable, supports international cooperation, and promotes fairness [4]. Other work leaves the debate over substantive exemptions to one side and proposes a procedural mechanism instead: a record of compliance covering the whole training process, supported by a three-stage model that shifts the burden of proof as evidence accumulates [2]. Case studies of Feilin v. Baidu, Tencent v. Yingxun, and the first AI text-to-image case trace a shift in Chinese practice from refusal to cautious recognition, and show that courts continue to decide each case on its facts and to insist that a human author stands behind the work [8]. Research on AI in chemistry makes a related point. Large, high-quality datasets are what make machine learning work, and scarcity, uneven quality, and bias between sources all reduce effectiveness, so that agreed reporting standards offer an institutional remedy [9]. A survey of memory-efficient training for Transformers notes that model size has grown roughly 240-fold every two years, concentrating the capacity to train large models in a few hands and widening the imbalance in access to data [10].

2.4 Assessment and the Approach Taken Here

The literature has advanced both the institutional design of copyright governance and the technical analysis of security risk, but two gaps remain. The two problems are usually treated separately, with little attempt to bring them into a single framework, and case studies tend to stay within one jurisdiction, so that Chinese and American decisions are rarely compared directly. This paper takes the comparison of leading cases as its point of departure and brings copyright and security together in one analysis.

3. METHOD

The study uses three standard methods of social research: case analysis, which supplies the material; comparison, which supplies the perspective; and normative analysis, which supplies the argument.

3.1 Case Analysis

The cases chosen are the leading copyright disputes over training data in the two jurisdictions. In the United States they are Anthropic, Meta, and *The New York Times v. OpenAI*; in China they are *Feilin v. Baidu*, *Tencent v. Yingxun*, the first AI text-to-image case, and *Shanghai Xinchuanghua v. Hangzhou Jellyfish Intelligence*. Facts, issues, and holdings are examined in order to show how courts have responded to the problem of training data.

3.2 Comparison

The cases, and the governance models behind them, are compared on three axes: the path the courts take, the standard applied in fair use analysis, and the way rights are allocated. The comparison identifies what the two systems share and where they part, and indicates what China might take from the American experience.

3.3 Normative Analysis

On the basis of that comparison, the paper argues for particular institutional designs, among them a locally adapted fair use provision, a mechanism for proving process compliance, and a system of standards, and offers recommendations intended to be both workable and new.

4. ANALYSIS

4.1 A Comparison of Chinese and American Cases

4.1.1 The United States: Fair Use Expanding and Contracting

American courts have addressed training data through the fair use doctrine. In the Anthropic case, decided in June 2025, the Northern District of California held that using copyrighted books to train the Claude model was fair use, the use being transformative and the data lawfully obtained [1]. In the Meta case that followed, the judge reached the same result on

a different ground, holding that Meta's unlicensed use of books by thirteen authors was not unlawful because the plaintiffs had not shown that the use caused them actual harm [1]. Together the two decisions suggest a tolerant attitude toward training and illustrate how much room the four-factor test, read together with transformative use, leaves to the individual judge.

The New York Times v. OpenAI points the other way. The complaint concerns ChatGPT's production of news summaries through retrieval-augmented generation, which diverts readers from the original articles and reduces traffic and advertising revenue. Because RAG requires complete copies of works to be held in a retrieval database, and because its output often serves the same function as the works retrieved, it substitutes for the original in the market and is unlikely to count as fair use [1]. Between Anthropic and The New York Times the case law thus moves in both directions at once, which shows how much the outcome depends on the facts of the individual case.

4.1.2 China: From Copyrightability of Output to Legality of Input

Chinese courts have moved from refusal to cautious recognition. Feilin v. Baidu (2019) held that copyright protects the work of natural persons, so that a report generated automatically by software was not a work [8]. Tencent v. Yingxun (2019) used the doctrine of works made for legal persons to leave room for protection, treating financial reports produced by the Dreamwriter program as protected literary works [8]. The first AI text-to-image case (2023) treated Stable Diffusion as an instrument: an image embodying the intellectual effort of the user was protected, but the model itself could not be an author [8]. The line of decisions shows two constant features, namely the insistence that a human being stands behind the work and the practice of deciding each case on its facts [8].

Infringement through training data has also reached the courts. In Shanghai Xinchuanghua Culture Co. v. Hangzhou Jellyfish Intelligence Technology Co., the Guangzhou Internet Court considered a request that Ultraman material protected by copyright be removed from the defendant's training set [5]. More generally, in The New York Times v. OpenAI and Tremblay v. OpenAI the issue has shifted from whether output qualifies as a work to where the line falls on the acquisition and use of input, which marks a movement of judicial attention from output to input [2].

4.1.3 Comparison and Implications

The two systems differ in ways that matter. American courts apply a four-factor test shaped by the idea of transformative use, and judges have wide discretion to find training transformative in the individual case; the cost is that outcomes are hard to predict [5]. Chinese courts insist on human authorship and decide case by case, and have only begun to address infringement through training data. Both face the same underlying difficulty: how to characterize the copying that training involves, and how to hold the balance between encouraging innovation and protecting rights holders. The American experience suggests that an open-ended fair use doctrine leaves room for technical development but leaves rights holders uncertain of their position. The Chinese experience suggests that cautious, incremental adjudication builds consensus but still needs legislative guidance on the prior question of whether training data was lawfully obtained.

Table 1 Leading Chinese and American cases on generative AI training data

Case	Jurisdiction / Year	Issue	Holding
Anthropic	United States / 2025	Whether using books to train a model is fair use	Use held transformative; fair use
The New York Times v. OpenAI	United States / 2023	Whether RAG summaries substitute for the original works	Market substitution at issue; fair use doubtful
Feilin v. Baidu	China / 2019	Whether an AI-generated report is a work	Natural-person authorship required; not a work
First AI text-to-image case	China / 2023	Ownership of copyright in an AI-generated image	Model treated as a tool; user holds copyright
Xinchuanghua v. Jellyfish Intelligence	China / 2024	Ultraman material present in the training set	Claimant sought removal of the training data

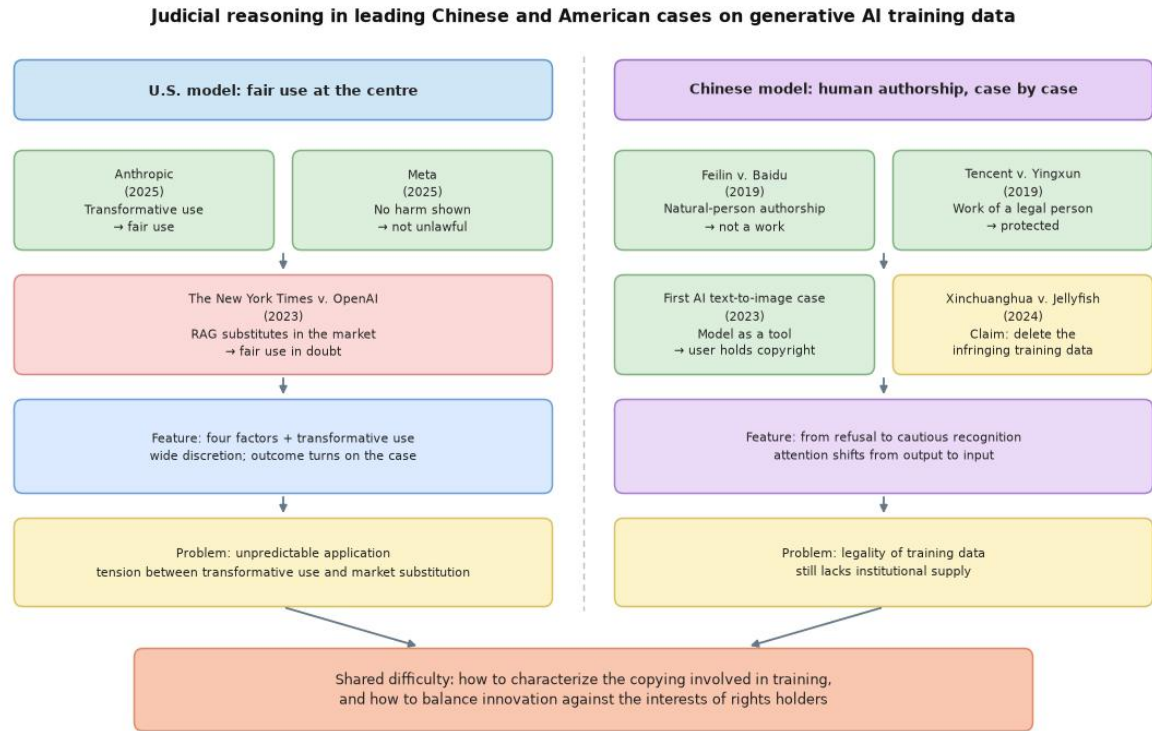


Figure 2 Judicial reasoning in leading Chinese and American cases on generative AI training data

4.2 The Compliance Problem in Copyright

4.2.1 Risk Across the Three Stages of Training

Training can be divided into data input, model learning, and model optimization, and the risk of infringement is not spread evenly across them [5]. It concentrates at input, where several rights are engaged at once: reproduction, through downloading and repeated copying; attribution, through the removal of authors' names and copyright notices; and integrity, through excerpting, alteration, and splicing. During learning, an encoder–decoder architecture breaks data down, abstracts it, and recombines it; the changes occur inside the machine and no work is communicated to the public, so the risk is smaller. Optimization involves only temporary storage of training data and the implicit retention of works in the form of parameters, and rarely raises questions of infringement. The unevenness of this distribution has a practical consequence: regulatory effort belongs at the input stage rather than across all three.

4.2.2 Why Fair Use Fits Badly

Fair use meets several obstacles in this setting. Training is a process of statistical optimization rather than the non-linear cognition of a human reader, so the claim that a model merely "learns" is weak [1]. The doctrine of non-expressive use would remove training from copyright altogether, but it does so by allowing developers to use works without payment, which transfers value in one direction [5]. Comparative experience points the same way. The American transformative use test is flexible but subjective, and its application is unpredictable; the Japanese provisions are confined to information and communication and permit copies to be communicated to the public, which cuts into the core of copyright; the European TDM exception allows rights holders to opt out, so firms cannot rely on the data being available [5]. Each has its own logic and its own defects, and none can be copied into Chinese law. RAG compounds the difficulty. It adds nothing transformative and substitutes for the original, so it is unlikely to benefit from any exception [1].

The substitution damage that training does to the potential market for a work can be measured with a market dilution coefficient, D [1]:

$$D = (S_AI / S_total) \times W \text{ (Formula 1)}$$

Here S_AI is the share of the market taken by AI-generated content, S_total the total size of the market in which the original work competes, and W a weight reflecting the harm. Under this model a value below 5 percent counts as no substantial effect and supports fair use; a value between 5 and 20 percent counts as minor and calls for modest compensation; at or above 20 percent the effect is substantial, the use is not fair, and it must stop, with damages payable [1]. The point of the formula is to give courts something measurable in place of the subjectivity of transformative use.

4.2.3 Licensing and Compensation

Licensing fails here for three reasons. Rights holders are hard to identify, since cleaning the data strips out metadata and leaves many orphan works; the corpus is enormous, ChatGPT-3 having been trained on about a terabyte; and policy favours technical progress and competition [5]. Collective management does not fill the gap, because it cannot cover every rights holder and every type of work, and because checking rights is incompatible with a practice of crawling

first, training later, and using whatever has been gathered. Two proposals have been put forward in response. One is a tiered licensing system, with direct licences for works of high value, collective licences for the mass of low-value works, and open licences for works deliberately released. The other is a compensation fund financed by a levy of 0.5 to 1 percent of revenue on AI firms earning more than 100 million yuan a year [1].

4.3 Security

4.3.1 Risks Across the Life Cycle

Coding 38 policy texts by the method of grounded theory identifies risks at six stages: collection, storage, annotation, computation, output, and deletion [6]. Collection carries risks of means, scope, and procedure, together with poisoning and poor-quality data; output carries risks of falsity, discrimination, social harm, infringement, and unaccountability. The map gives governance a structure to work with. Interaction with users makes collection less visible to the person concerned and leakage harder to detect, and inference carries the further risk that private items in the training set are recombined into output shown to someone who had no business seeing them [4]. Governance itself has three weaknesses: several regulators act without coordination, the rules available are too few and hard to apply, and administration dominates a system that lacks other forms of participation. Agile governance and wider participation are the usual answers [6].

The correspondence between the risks of each stage and the three-part governance system is shown in Figure 3.

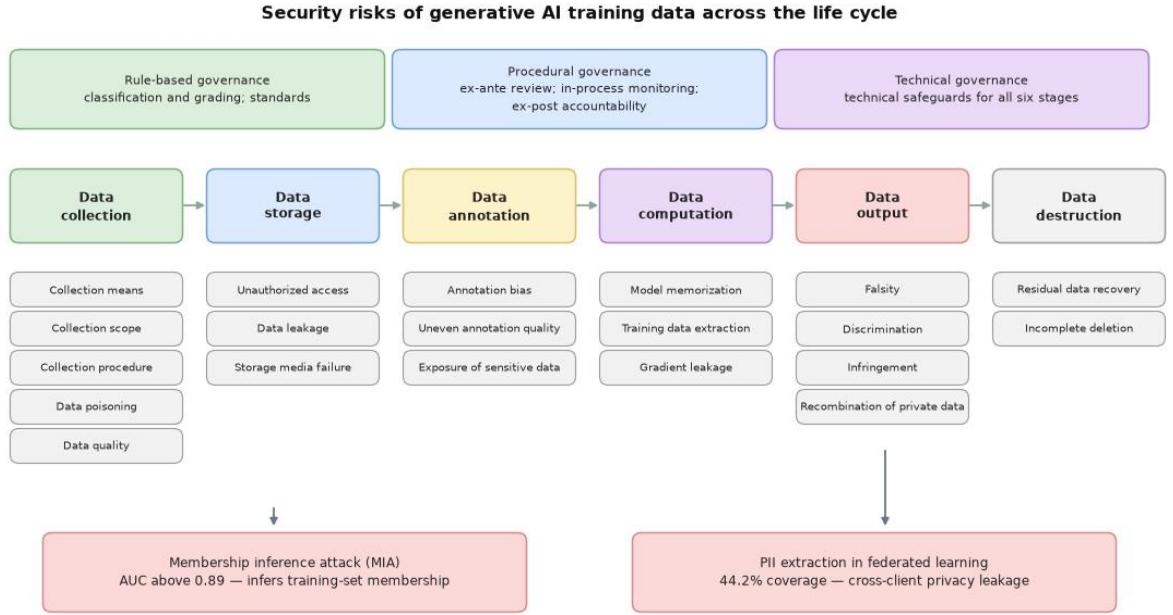


Figure 3 Security risks of generative AI training data across the life cycle

4.3.2 Memorization and Membership Inference

The membership inference attack is the main route by which training data loses its privacy. Networks with many parameters, tuned for maximum accuracy, tend to memorize particular items, especially atypical ones in the tail of the distribution, as a by-product of generalization, and the model then carries private information within itself [7]. The attack works because a model responds differently to data it has seen, generally with higher confidence and accuracy, so that querying the model is enough to tell whether an item was in its training set.

Memorization is inversely related to the entropy of the data, a relationship sometimes called the entropy–memorization law [7]:

$$M(z) = \alpha - \beta \cdot H(z) \text{ (Formula 2)}$$

Here $M(z)$ is the strength with which the model has memorized sample z , $H(z)$ the entropy of that sample, and α and β constants characteristic of the model. Low-entropy samples, which carry little information and occur rarely, are memorized more strongly and are correspondingly easier to identify [7]. Attack performance is measured by the area under the ROC curve:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}) \text{ (Formula 3)}$$

TPR is the rate at which member samples are correctly identified and FPR the rate at which non-members are mistaken for members. A value near 1 means the attack works; a value near 0.5 means it does no better than chance. Attacks based on differences between subgroups have reached an AUC above 0.89 on CIFAR-10, and an attack on video data reached 97.01 percent under black-box conditions, which suggests the threat is practical rather than theoretical [7]. Its existence also has a legal consequence. A model can no longer be treated simply as software, since it may contain personal data that can be recovered, and it is regulated accordingly [7].

4.3.3 Federated Learning

Federated learning is often presented as privacy-preserving, but it leaks as well. The FedLLM-Legal benchmark and the FedLLM-Extract attack show a semi-honest client building a set of frequent prefixes from its own data, sampling the global model repeatedly, and estimating how much personal information sits in the data of other clients [3]. Coverage is measured as:

$$CR = | PII_extracted \cap PII_target | / | PII_target | \text{ (Formula 4)}$$

$PII_extracted$ is what the attacker recovers and PII_target the personal information held by the client under attack. Across five tasks the attack reached a coverage of 44.2 percent, and because it can be repeated in every round of training it recovers more than twice what a single attempt yields. The study annotated personal information in seven categories and thirty-six subcategories, with an accuracy of 89 percent [3]. Even a paradigm designed for privacy therefore leaves sensitive material open to extraction, which argues for combining technical safeguards with rules.

4.3.4 Technical Safeguards

Vp3CNN, which combines VOLE-ZKP with secret sharing, performs convolutional inference across three parties with both privacy and verifiability. Input is split by two-out-of-two additive secret sharing between two servers assumed not to collude; each computes on its share without learning the underlying data, and the correctness of the result is established by a zero-knowledge proof based on vector oblivious linear evaluation [11]. On MNIST the scheme reaches 97.8 percent accuracy and verifies convolution four to five times faster than earlier methods, so privacy and verifiability need not conflict. Adopting such a scheme also builds

confidence in automated systems and helps firms meet regulatory requirements on security and privacy [11]. On the defensive side, the GAN-SOFT approach lowers attack AUC to about 0.58 while costing between 1 and 3 percent in accuracy, and the Entro-DWE ensemble brings the success rate on the WikiMIA benchmark close to chance [7]. The object of defence has shifted: rather than preventing information from being read, it now aims at removing memorization from the model [7].

4.4 A Governance Framework

The argument so far points to four elements working together: procedural proof alongside substantive rules, a fair use provision adapted to Chinese conditions, standards as a lever, and agile governance shared among participants (Figure 1). Copyright and security are treated as parts of one problem.

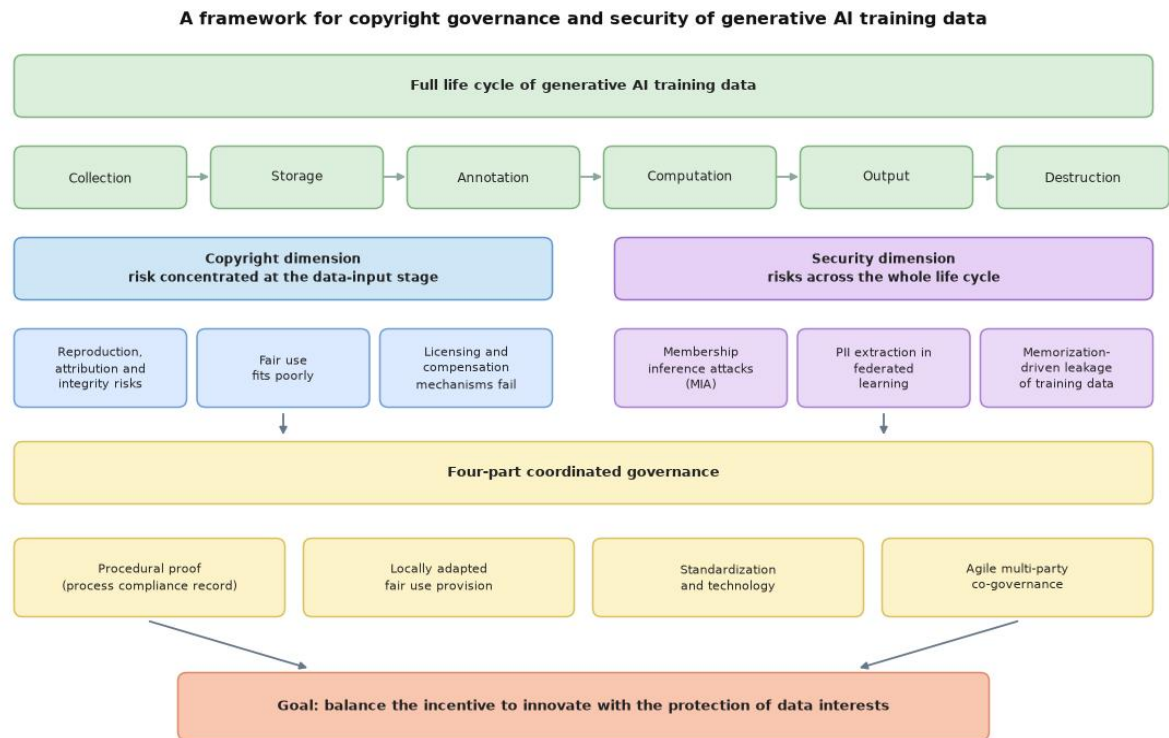


Figure 1 A framework for copyright governance and security of generative AI training data

4.4.1 Proving Compliance

The immediate difficulty is that a developer has no way to show that training was lawful. Rather than continue to argue about the conditions of substantive exemption, the law might ask for proof of procedure, in the form of a record of compliance kept through the whole process of training [2]. The "judicially credible compliance record package" proposed for this purpose has three layers: provenance and preprocessing of the input data; documentation of the training objective together with evidence that the computing environment was isolated; and monitoring of output through watermarks and similarity reports. It is governed by four principles, namely to collect only what is necessary, to keep a record throughout, to verify by cross-checking, and to preserve evidence with a third party. A companion model allocates the burden of proof in three stages: once the developer produces a complete package, lawfulness is presumed and the burden moves to the rights holder, who can rebut it only by showing that the package was systematically falsified or that a shadow system existed outside what was recorded [2]. Requirements of this kind write the law into the development process. Regulators keep a point of control, firms get a clear standard to meet, and neither absolute exemption nor prohibition is required.

4.4.2 A Fair Use Provision for China

Substantive reform would follow the typological approach of Japan and the European Union, adding an exception for the use of lawfully accessed works to the extent necessary for the analysis of data and information, and specifying the elements of the three-step test, namely purpose, manner, and consequence of use, in the factorial style of the United States [5]. The policy of inclusive and prudent regulation suggests three routes working at once: legislation opening the residual clause of the Copyright Law to a non-enjoyment use exception, adjudication developing tiered tests and reversing the burden of proof, and industry promoting technical empowerment and compliance certification [1]. Fair use also needs support from two other institutions. One is disclosure of the provenance, ownership status, and licensing of datasets and of any synthetic data used; the other is compensation, assessed by the top-down method used for standard-essential patents, at somewhere between 5 and 10 percent of AI revenue [5]. Fair use marks the boundary of permitted conduct; disclosure and compensation adjust the distribution of benefits within it.

4.4.3 Standards and Technology

Standardization has been called the first move against the risks of generative AI, since it makes technology more controllable, supports international cooperation, and promotes fairness [4]. In the governance of training data, standards for provenance, verification of rights, quality assessment, and security audit would cover the data through its whole life [6]. The experience of AI in chemistry is instructive here as well: machine learning depends on large, high-quality datasets, and scarcity, uneven quality, and bias between sources all reduce what it can do, so agreed reporting standards serve as an institutional remedy [9]. Provenance labels and declarations of rights, set out in a standard form, are the infrastructure on which compliance depends. Grading provenance compliance, L1 for the basics and L3 for full traceability, would also allow risk to be regulated in proportion: high-risk applications would be held to L3, low-risk ones to the reduced requirements of L1 [2].

4.4.4 Agile Governance and Participation

Three weaknesses in the present system have already been noted: regulators acting without coordination, too few rules and too hard to apply them, and an administrative approach unsupported by other participants. Agile governance, with several parties involved, is the response usually proposed [6]. The concentration of capacity matters as well. Since model size has grown roughly 240-fold every two years, the ability to train large models belongs to a few organizations, and the imbalance in access to data widens accordingly [10]. Governance therefore has to look beyond copyright and privacy to the structure of the industry. Membership inference attacks can be turned to good use here: as an automated privacy check, they give firms a way to measure risk before a model is released, which is a case of technique serving regulation [7]. Regulatory sandboxes give supervisors room to observe and to adjust, and soft law, being more flexible, can be revised as the technology changes [6].

5. LIMITATIONS AND PROSPECTS

The study works from published cases, policy documents, and secondary literature. No original data were collected, no estimates were computed, and no attacks were reproduced. Figures quoted here for market size, attack performance, coverage, and model accuracy are taken from the studies cited and should not be read as findings that this paper has verified. Later

work could publish the data, the coding rules for cases, and the scripts used, so that the claims made here can be checked.

6. CONCLUSIONS

Comparison of the leading Chinese and American cases supports four conclusions.

First, the two jurisdictions have taken different paths. American courts, applying fair use and the idea of transformative use, accepted training in Anthropic and Meta but treated the substitution caused by RAG in *The New York Times v. OpenAI* with suspicion, so that the case law pulls in both directions at once. Chinese courts, insisting on human authorship and deciding case by case, moved from *Feilin v. Baidu* to the first AI text-to-image case and from refusal to cautious recognition, with attention shifting from the copyrightability of output to the lawfulness of input. Both systems still have to settle how the copying involved in training should be characterized and how innovation and protection are to be balanced.

Second, the compliance problem has three sources: the uneven distribution of risk across the stages of training, the poor fit of fair use, and the failure of licensing and compensation. Risk is highest at input, and because non-expressive use transfers value in one direction while licensing no longer functions, rules on fair use have to be accompanied by disclosure and compensation.

Third, security risks extend across the whole life cycle. Membership inference exploits memorization to identify members of the training set, reaching an AUC above 0.89 in the experiments cited, and extraction attacks in federated learning reached 44.2 percent coverage in those experiments, so that even a paradigm designed for privacy leaks. Schemes such as Vp3CNN and defences such as GAN-SOFT show that protection is moving from preventing access to removing memorization, which gives regulation something to build on.

Fourth, governance needs the four elements described above: procedural proof to establish lawfulness, a fair use provision adapted to local conditions, standards as the infrastructure of compliance, and agile governance to keep pace with technical change. The paper offers these as a framework for balancing innovation in generative AI against the interests of those whose data it uses. How well each element works, and how well they work together, remains to be tested against evidence.

REFERENCES

- HE Junhui & GAO Chengjun. (2026). Analysis of the fair use rule in generative AI copyright: From the perspective of the U.S. Copyright Office report Copyright and Artificial Intelligence. *Future and Development*, 50(5), 20-28. (in Chinese)
- HU Shensong & WANG Ruoyu. Process compliance and procedural safeguards: Functional reshaping and rule construction of AIGC training-data provenance. *Journal of Hangzhou Dianzi University (Social Sciences)*, 1-13. (in Chinese)
- HU Yingqi. (2025). Benchmark construction for federated large models in legal scenarios and research on privacy attacks (Master's thesis, Harbin Institute of Technology). <https://doi.org/10.27061/d.cnki.ghgdu.2025.002642>. (in Chinese)
- RU Peng, SU Jun, HAN Zhihong & SHU Quanfeng. (2025). Intelligence leading the future: Social impacts of generative artificial intelligence and governance through standards. *E-Government*, (1), 2-14. <https://doi.org/10.16582/j.cnki.dzzw.2025.01.001>. (in Chinese)
- XIE Liwei & PAN Sheng. The copyright compliance dilemma of generative AI data training and paths of governance. *Journal of Shandong University of Science and Technology (Social Sciences)*, 1-12. <https://doi.org/10.16452/j.cnki.sdkjsk.20260617.002>. (in Chinese)
- XU Wei & HE Ye. (2024). The governance system for generative AI data security risks and its optimization: A grounded analysis of 38 policy texts. *E-Government*, (10), 42-58. <https://doi.org/10.16582/j.cnki.dzzw.2024.10.005>. (in Chinese)
- YANG Jia. Membership inference attacks on machine learning models and mechanisms of privacy protection [D]. Supervisors: ZHANG Xing; MIAO Chunyu. Liaoning University of Technology, 2026. (in Chinese)
- YUAN Xiuting & YAN Chi. Some typical cases on copyright protection of AI-generated content. *Journal of Ocean University of China (Social Sciences)*, 1-11. (in Chinese)
- DING Hu, HUA Pengxiang & HUANG Zhen. (2026). Survey on recent progress of AI for chemistry: methods, applications, and opportunities. *Frontiers of Computer Science*, 20(11).

TIAN Kaiyuan, QIAO Linbo, LIU Baihui, JIANG Gongqingjian, LI Shanshan & LI Dongsheng. (2026). A survey on memory-efficient transformer-based model training in AI for science. *Frontiers of Computer Science*, 20(11).

Shengnan Zhao, Kuiheng Sun, Chuan Zhao, Bendong Jiang, Zhenxiang Chen & Qiuliang Xu. (2025). Vp3CNN: A verifiable privacy-preserving three-party scheme for convolutional neural network inference. *Chinese Journal of Electronics*, 34(5), 1427-1437.