COPYRIGHT ALIGNMENT FOR GENERATIVE AI: DESIGNING A DUAL-PHASE DUTY-OF-CARE REGIME IN CHINA

This paper examines how Chinese copyright law can reconcile authors' rights with generative-AI innovation through a dual-phase duty-of-care regime. It proposes recognising AI training as a form of fair use or a copyright exception conditioned on "copyright alignment" during training – requiring lawful data sourcing, dataset governance, and anti-memorisation safeguards – and establishing proportionate duties of care during deployment, including filtering, watermarking, and "notice and necessary measures" mechanisms under Arts 1195 and 1197 of the Civil Code of the People's Republic of China. "Necessary measures" encompass deleting infringing content, blocking prompts, and optimising models to prevent recurrence. Drawing on Chinese cases and comparative regimes, this paper argues that technically grounded obligations can mitigate infringement risks while fostering innovation and responsible AI governance in China.

Jianfeng **CAO**
PhD in Law (University of International Business and Economics); Research Fellow, Internet Law Research Center, South China University of Technology.

Jiahui **CHEN**
LLM (Jinan University); PhD Candidate, Queen Mary University of London.

I. Introduction

1 In recent years, large language models and other generative artificial intelligence ("AI") systems have advanced at a remarkable pace, pushing the field toward ambitious goals of artificial general intelligence and even superintelligence. At the same time, they have triggered novel and complex copyright challenges worldwide, including in China. These systems, which range from large language models to image and audio generators, including music composition tools, are typically trained on internet-scale datasets scraped from the web, often including copyrighted

texts, images, audio, and videos.[1] Once deployed, they can generate a wide variety of content. This dynamic gives rise to two major friction points for generative AI: (a) the use of copyrighted materials in AI model training without authorisation or licensing from rights holders;[2] and (b) deployment risks when user prompts or model behaviour result in infringing or derivative outputs.[3]

2 Under traditional copyright law, both stages could implicate content creators' rights, yet the application of old rules to generative AI's new capabilities is fraught with uncertainty. Chinese copyright law must address two central issues raised by generative AI: the legality of using copyrighted materials as training inputs, including whether such use may qualify as fair use, and the allocation of liability when AI outputs infringe or adapt protected works. Recent developments in China's legal landscape, including new administrative regulations[4] and high-profile judicial rulings,[5] indicate a potential trajectory toward establishing a "copyright duty of care" for generative AI providers, designed to mediate the tension between technological innovation and copyright protection.

3 China has moved swiftly in recent years to confront the copyright issues raised by generative AI. While the 2020 amendment to the Copyright Law of the People's Republic of China[6] represented a critical update, it did not expressly resolve the status of AI training or AI-generated works under the statute. Instead, subsequent developments have primarily come through administrative regulations and judicial cases.

---

1 Priyanka Gupta *et al*, "Generative AI: A Systematic Review Using Topic Modelling Techniques" (2024) 8(2) *Data and Information Management* 100066.

2 J Quang, "Does Training AI Violate Copyright Law?" (2021) 36 *Berkeley Technology Law Journal* 1414.

3 Eleonora Rosati, "Infringing AI: Liability for AI-Generated Outputs Under International, EU, and UK Copyright Law" (2025) 16(2) *European Journal of Risk Regulation* 603 at 619–621.

4 Interim Measures for the Administration of Generative Artificial Intelligence Services (生成式人工智能服务管理暂行办法) (effective 15 August 2023) (China); Administrative Provisions on Deep Synthesis of Internet Information Services (互联网信息服务深度合成管理规定) (effective 10 January 2023) (China); Measures for the Labelling of AI-Generated Content (人工智能生成合成内容标识办法) (effective 1 September 2025) (China).

5 *Shanghai Character License Administrative Co Ltd v Guangzhou Nianguang Network Technology Co Ltd* (Guangzhou Internet Court) (2024) Yue 0192 Min Chu (Civil preliminary trial) No 113 ((2024) 粤0192民初113号); *Shanghai Character License Administrative Co Ltd v Hangzhou Jellyfish Intelligent Technology Co Ltd* (Hangzhou Internet Court) (2024) Zhejiang 0192 Min Chu (Civil preliminary trial) No 1587 ((2024)浙0192民初1587号); (Hangzhou Intermediate Court) (2024) Zhejiang 01 Civil Min Zhong (Civil final instance) No 10332 ((2024)浙01民终10332号).

6 (中华人民共和国著作权法) (effective 1 June 2021) (China).

4 In July 2023, Chinese authorities issued the Interim Measures for the Management of Generative AI Services,[7] making China the first country to adopt regulations specifically targeting generative AI. These measures codify the principle of respecting intellectual property rights in AI development and deployment by requiring service providers to ensure that their training data are lawfully sourced and do not infringe upon the copyrights of others. Complementing this framework, the Administrative Provisions on Deep Synthesis of Internet Information Services[8] and subsequent Measures for the Labelling of AI-Generated Content[9] impose additional obligations concerning disclosure, traceability, and labelling. Taken together, these instruments signal a regulatory shift toward proactive compliance and accountability for providers of generative AI. At the same time, however, they leave unresolved the central legal question of whether the use of copyrighted works as training data, particularly for commercial AI models, constitutes permissible use or infringement under Chinese copyright law. This unresolved issue now lies at the heart of ongoing litigation, such as *iQIYI v MiniMax*[10] and the Beijing Internet Court's AI-painting cases.[11]

5 Meanwhile, Chinese courts have started to address disputes involving generative AI, with several landmark rulings in 2024 clarifying copyright obligations and the duty of care of AI providers. For instance, the Guangzhou Internet Court held an AI company liable for infringing images generated by its model that closely resembled the famous Japanese character, Ultraman, signalling that generative AI service providers may bear responsibility for infringing outputs.[12] Later that year, the Hangzhou Internet Court, upheld on appeal by the Hangzhou Intermediate Court, further examined an AI provider's liability for infringing outputs generated through user prompts, while also considering whether users' personal training activities, such as fine-tuning an AI model with copyrighted images, could qualify as

7 (生成式人工智能服务管理暂行办法) (effective 15 August 2023) (China).
8 (互联网信息服务深度合成管理规定) (effective 10 January 2023) (China).
9 (人工智能生成合成内容标识办法) (effective 1 September 2025) (China).
10 MLex correspondent, "iQIYI Sues Chinese AI Startup MiniMax for Copyright Infringement in Landmark Case", *LexisNexis* (3 January 2025) <https://www.lexisnexis.co.uk/legal/news/iqiyi-sues-chinese-ai-startup-minimax-for-copyright-infringement-in-landmark-case> (accessed 19 July 2025).
11 Berry Wang & Jessie Yeung, "Chinese Artists Boycott Big Social Media Platform Over AI-Generated Images", *CNN Business* (28 September 2023) <https://edition.cnn.com/2023/09/28/tech/chinese-artists-boycott-ai-generator-intl-hnk> (accessed 19 July 2025).
12 *Shanghai Character License Administrative Co Ltd v Guangzhou Nianguang Network Technology Co Ltd* (Guangzhou Internet Court) (2024) Yue 0192 Min Chu (Civil preliminary trial) No 113 ((2024) 粤0192民初113号).

fair use under Chinese copyright law.[13] However, the court did not definitively resolve the central question of whether the use of copyrighted works as training data for commercial generative AI models constitutes fair use or infringement. This unresolved issue lies at the heart of several other ongoing cases: (a) *iQIYI v MiniMax*, in which the video platform alleged that MiniMax trained its models using iQIYI's copyrighted video materials without authorisation and; (b) four cases before the Beijing Internet Court, where illustrators sued the developers and operators of an AI-painting software for using their artworks as training data without permission and for commercial purposes. The outcomes of these pending cases are expected to exert a profound influence on the development of generative AI, particularly given the absence of an explicit provision in Chinese copyright law regarding the legal status of using copyrighted materials for AI training.

6 Against this backdrop, this paper examines the emerging notion of a "copyright duty of care" for generative AI providers and its implications within the Chinese copyright regime. This inquiry is guided by two central research questions: (a) What obligations do generative AI providers bear in preventing copyright infringement at both the training and deployment stages of AI development? (b) How should Chinese copyright law conceptualise and regulate such obligations across the AI lifecycle, from dataset preparation and model training to user-facing content generation? To address these questions, the paper proceeds as follows. Part II analyses recent judicial decisions and theoretical foundations, identifying the contours of a nascent dual-stage framework of generative AI providers' copyright duty of care. Part III then explores the scope of obligations in the training phase, assessing whether the use of copyrighted materials for AI training may fall within the ambit of fair use, and considering what forms of "copyright alignment" providers should implement to mitigate copyright infringement risks. Part IV examines the deployment phase, focusing on providers' liability for infringing outputs and preventive or remedial measures, such as filatering, monitoring, and take-down mechanisms, that may be expected of them. Finally, Part V offers policy-oriented reflections on balancing rights protection with AI innovation, suggesting that a clearly articulated duty of care concerning copyright protection could form a cornerstone of responsible AI governance in China, and concludes with policy recommendations to guide future regulatory development.

13 *Shanghai Character License Administrative Co Ltd v Hangzhou Jellyfish Intelligent Technology Co Ltd* (Hangzhou Internet Court) (2024) Zhejiang 0192 Min Chu (Civil preliminary trial) No 1587 ((2024)浙0192民初1587号); (Hangzhou Intermediate Court) (2024) Zhejiang 01 Min Zhong (Civil final trial) No 10332 ((2024) 浙01民终10332号).

II. Case study and theoretical review

A. Decisions of Chinese courts

7 The Ultraman copyright infringement dispute adjudicated by the Guangzhou Internet Court[14] ("*Guangzhou Ultraman*") and the related copyright infringement and unfair competition dispute adjudicated at first instance by the Hangzhou Internet Court and on appeal by the Hangzhou Intermediate Court[15] ("*Hangzhou Ultraman*") together mark the world's first judicial efforts to deliver a coherent response to copyright infringement issues arising from generative AI service providers. These cases do not only set essential precedents in China but also contribute to the global discourse on copyright liability in the age of generative AI.

(1) Guangzhou Ultraman

8 In *Guangzhou Ultraman*, the plaintiff, a Shanghai cultural company, held the exclusive license for the copyrights of the Ultraman series in mainland China. The defendant, a Guangzhou internet technology company, operated a generative AI service named "Tab". This platform, which connected *via* an Application Programming Interface to a third-party service provider, offered a paid AI-powered drawing function. Users, by entering text prompts such as "Generate Ultraman" or "Ultraman with long hair," could prompt the system to generate and download images that were found to be substantially like the Ultraman character. The plaintiff alleged that this activity infringed its rights of reproduction, adaptation, and communication to the public through an information network.

9 The court's judgment centred on three main legal questions: (a) the determination of copyright infringement; (b) the nature of the generative AI service provider's liability; and (c) the specific remedies to be implemented. Regarding copyright infringement, the court found that the generated images were substantially similar to the original artistic expression of the Ultraman character by applying the "substantial similarity + access" test, which is frequently referred

14 *Shanghai Character License Administrative Co Ltd v Guangzhou Nianguang Network Technology Co Ltd* (Guangzhou Internet Court) (2024) Yue 0192 Min Chu (Civil preliminary trial) No 113 ((2024) 粤0192民初113号).

15 *Shanghai Character License Administrative Co Ltd v Hangzhou Jellyfish Intelligent Technology Co Ltd* (Hangzhou Internet Court) (2024) Zhejiang 0192 Min Chu (Civil preliminary trial) No 1587 ((2024) 浙0192民初1587号); (Hangzhou Intermediate Court) (2024) Zhejiang 01 Min Zhong (Civil final trial) No 10332 ((2024) 浙01民终10332号).

to as "potential exposure" in the context of online services. Given the character's immense fame and broad public exposure in China, the court presumed the defendant's access to the copyrighted works. The court held that the defendant's service, by generating these infringing images, directly infringed the plaintiff's rights of reproduction and adaptation.

10 With respect to the determination of liability for the generative AI service provider, the court found that the defendant was a direct infringer, not merely an accessory. The key factors in this determination were the defendant's actual control over the AI model's outputs and its failure to exercise reasonable care and diligence. The court explicitly rejected the defendant's arguments, including assertions that user prompts were the direct cause of the infringement and that a third party provided the underlying AI model. The court reasoned that the defendant, as the operator of the service, had the technical ability and legal obligation to control the output.

11 Furthermore, the court referenced the Interim Measures for the Administration of Generative AI Services. The defendant's failure to implement required safeguards, such as a complaint-and-reporting mechanism, explicit copyright risk warnings, and prominent watermarks on generated content, demonstrated a lack of reasonable care, solidifying the finding of subjective fault and supporting the imposition of direct infringement liability.

12 Regarding liability and remedies, the court ordered the defendant to cease infringing activity immediately. This included implementing technical measures to prevent its generative AI service from generating images substantially like the Ultraman character when users enter related prompts. The court specified that the required standard should be that the service is functionally incapable of generating infringing content. However, the court denied the plaintiff's request for the defendant to remove relevant content from its training dataset because the defendant itself did not conduct the model training. This distinction highlights the court's nuanced approach to liability, focusing on the actions and responsibilities of the platform operator rather than the underlying model provider.

(2) Hangzhou Ultraman

13 In *Hangzhou Ultraman*, the plaintiff, a Shanghai cultural company, held the exclusive copyright license for the Ultraman series in mainland China. The defendant, a Hangzhou intelligent technology company, operated a generative AI platform providing users with access to a third party's Stable Diffusion model. The platform offered various services, including text-to-image generation and online model

training, both as free and paid features. The platform's homepage and "Recommended" sections featured numerous AI-generated images and low-rank adaptation ("LoRA") models related to the Ultraman works. An investigation revealed that the infringing LoRA models were created by users who uploaded Ultraman images, trained the models, and subsequently generated derivative works. The plaintiff argued that the defendant's actions infringed its exclusive right of information network communication.

14 The Hangzhou Internet Court's ruling primarily focused on the legal issues analogous to those in *Guangzhou Ultraman*. Regarding infringement, the court found that the defendant, as a service platform, did not directly create the infringing content. Instead, it facilitated the process by providing users with the tools and platform to upload, train, and generate infringing content. This process, which enabled the online dissemination of infringing images and models, was found to violate the plaintiff's exclusive right of information network communication.

15 As for the generative AI service provider, the court meticulously analysed its liability, first ruling out direct infringement. The court applied a two-part "objective behaviour + subjective intent" criterion. According to the judgment, "objective behaviour" refers to the fact that the defendant did not directly perform the infringing acts, such as uploading images or entering prompts. In contrast, the subjective intent concerns the absence of evidence of shared intent between the defendant and its users to create infringing works jointly. Based on this analysis, the court precluded direct infringement liability.

16 The court then proceeded to determine secondary liability by applying a "reasonable person standard in the same industry" test to assess the defendant's duty of care. This standard evaluates whether a platform operator exercised sufficient care to prevent infringement, considering both the foreseeability of harm and the technical feasibility of prevention. The court assessed the defendant's conduct from several perspectives, including "nature of the service", which requires the defendant, as a generative AI service provider, to have sufficient understanding of the content generated on its platform and to exercise a corresponding duty of care. The second is "fame of the work", referring to the high popularity of Ultraman, which made infringement highly foreseeable. The third is "foreseeability of harm", which means the defendant should have anticipated the possibility of infringement due to its services. The fourth is "profit model", where the defendant's direct economic benefits from the service created a heightened duty of care. Finally, with regard to "preventive measures", the defendant was found to have possessed the technical capacity to implement reasonable measures to prevent infringement but failed to do so.

17 Considering these findings, the court concluded that the defendant had not exercised reasonable care and was at fault, thereby constituting contributory infringement (also known as secondary or indirect infringement).

18 The court ordered the defendant to cease its infringing activities immediately. The required remedies included the deletion of infringing content, which meant that the defendant was ordered to remove all infringing images and LoRA models that incorporated the original Ultraman features and were publicly available on the platform. Additionally, the defendant was required to cease publication and application services for the relevant Ultraman LoRA models. Moreover, the defendant was mandated to adopt effective measures to prevent future infringement.

19 This verdict highlights a key legal distinction between providing a tool for infringement and directly participating in it, establishing a precedent for assessing the secondary liability of AI service platforms based on their role and duty of care.

B. Examination of courts' decisions

20 While both the Guangzhou and Hangzhou Ultraman cases are landmark rulings on generative AI and copyright in China, they demonstrate fundamentally different judicial approaches to determining an AI service provider's liability. The core distinction lies in whether the court viewed the AI platform as a direct infringer or a facilitator of infringement.

21 The Guangzhou Court's decision was based on the premise that the defendant's AI platform was a direct party to the infringement. This finding was rooted in the technical process of the service provided. The platform operated a text-to-image service where the underlying AI model had been pre-trained on a massive dataset that included copyrighted works like Ultraman.

22 When a user entered a prompt such as "generate Ultraman", the AI model used its pre-existing "knowledge" derived from the stored training data to generate an infringing image. The court ruled that this act of generating the image constituted reproduction (through the use of stored features), and adaptation (creation of a new, derivative work). The platform itself was seen as the agent performing the infringing acts. But the court did not answer whether this act also constituted communication through information networks (or communication to the public through information networks).

23 Further, the court's reasoning could be understood through the analogy of a factory. The defendant's platform could be likened to a factory pre-loaded with moulds for a copyrighted product. When a user places an order, the factory's automated system produces the infringing item. In this scenario, the factory owner, *ie*, the platform, is the direct infringer because it is the one manufacturing the infringing product. The user is merely the customer.

24 The court also found the platform liable for failing to exercise a reasonable duty of care. This was not just a failure to respond to a notice, but a systemic failure to comply with regulations. The defendant had not implemented a complaint-and-reporting mechanism, provided adequate risk warnings, or applied prominent watermarks to the outputs, demonstrating an apparent subjective fault.

25 Conversely, the Hangzhou Court's verdict established the defendant's role as a contributory infringer. The technical process on this platform was fundamentally different: it provided the tools and infrastructure for users to create and share their own AI models.

26 The defendant's platform offered a basic AI model and allowed users to upload their own images to train LoRA models: the small models that fine-tune the base AI model. The users themselves uploaded the Ultraman images and trained the infringing models. The platform did not independently store or use the Ultraman images for its core training.

27 In this case, the platform's role was likened to a service provider offering tools and a space for users to act. The users performed the infringing act of creating and disseminating the infringing models and images. The platform's liability arose from its knowledge of the infringing activity and its failure to take reasonable measures to stop such activities. This fits the legal definition of aiding and abetting an infringement.

28 Accordingly, the court found that the defendant's platform prominently displayed infringing Ultraman models and images, making the infringement obvious. The platform also profited from these user-generated models. Based on this, the court concluded that the defendant knew or should have known of the infringement but failed to act; a finding that is consistent with the "knowledge" rule under Chinese copyright law.

29 The verdicts in these two cases provide critical guidance for the legal landscape of generative AI by establishing a clear distinction based on the platform's direct and indirect involvement in the infringement process.

30 The most significant distinction is the finding of direct infringement in *Guangzhou Ultraman versus* the finding of contributory infringement in *Hangzhou Ultraman*. This reflects a fundamental difference in how the courts characterised the defendant's role, as a principal actor in the Guangzhou case *versus* a facilitator in the Hangzhou case.

31 The Guangzhou Court held the platform responsible for infringement at the output/deployment stage, where the generated content was a direct result of its pre-trained model. The Hangzhou Court, conversely, focused on the dissemination stage, holding the platform liable for failing to monitor and manage user-uploaded infringing content that was highly visible.

32 Further, the Guangzhou Court's verdict was based on a direct application of copyright infringement principles combined with the non-compliance of new AI regulations. In contrast, the Hangzhou Court relied on traditional secondary liability doctrines and the "knowledge" rule to address the unique challenges of AI-assisted infringement.

33 Together, these cases underscore that AI service providers must exercise a significant duty of care. However, the specific nature of that duty and resulting liabilities will be determined by their particular business model and direct involvement in the creation and dissemination of infringing content.

C. Comparative theoretical review

34 The emergence of generative AI has presented novel challenges to established tort and copyright law frameworks, as demonstrated by two landmark Ultraman copyright infringement cases adjudicated by the Chinese Internet Courts in Guangzhou and Hangzhou.

35 The divergent verdicts of the two Ultraman cases reflect the conceptual challenges posed by generative AI platforms, which occupy an ambiguous position within existing legal taxonomies. Unlike traditional internet service providers that merely host or transmit user-generated content, generative AI platforms actively process user prompts to create new content through algorithmic generation. This dual nature, combining elements of passive infrastructure provision with active content creation, distinguishes these platforms from conventional categories of internet content providers or service providers recognised under current legal frameworks. The platforms' role in content generation is neither purely passive nor entirely autonomous, as they respond to

user inputs while employing proprietary algorithms and training data to produce outputs.

36 Legal complexity arises from determining whether generative AI platforms directly perform actions controlled by exclusive copyright rights, constituting direct infringement, or whether liability should be assessed based on their facilitation of user-initiated infringing activities. As technological innovation continues to outpace legal adaptation, courts and legislators face the fundamental question of whether existing intellectual property and tort law doctrines can be adequately extended to address these novel scenarios, or whether comprehensive regulatory frameworks specifically designed for AI are necessary to provide legal certainty and appropriate protection for both creators' rights and technological innovation.

37 In contrast to the approaches taken by Chinese courts, the current landscape of US litigation involving generative AI and copyright infringement presents a more complex and evolving jurisprudential framework. The fundamental challenge lies in applying traditional Digital Millenium Copyright Act[16] ("DMCA") safe harbour provisions, which were designed for passive intermediaries, to generative AI platforms that actively create content.[17] While some early analyses suggested that generative AI platforms might benefit from safe harbour protections, qualification for such protection requires platforms to demonstrate they lack knowledge of infringement and implement effective take-down procedures once notified.[18] Recent judicial developments indicate a more nuanced approach to AI copyright issues. In *Thomson Reuters Enterprise Centre GmbH v Ross Intelligence Inc*,[19] the Delaware Federal Court issued the first significant ruling concerning the use of copyrighted material to train AI systems. Meanwhile, another federal judge ruled that AI companies feeding copyright-protected works into their models without permission or payment generally violate the law. These decisions suggest that US courts are not uniformly applying safe harbour protections to AI platforms, particularly regarding the training phase of AI development.

38 Within the European Union ("EU"), the regulatory approach to generative AI copyright issues has been comprehensively addressed through a study released by the European Union Intellectual Property

16 Copyright Act 17 USC (US) § 512.
17 Michael P Goodyear, "Infringing Information Architectures" (2025) 58(4) *UC Davis Law Review* 1959 at 1960.
18 *Andersen v Stability AI Ltd* No 3:23-cv-00201 (13 January 2023, ND Cal) (US); *Kadrey v Meta Platforms Inc* No 23-cv-03417 VC (20 November 2023, ND Cal) (US).
19 694 F Supp 3d 467 (D Del, 2023).

Office ("EUIPO").[20] This EUIPO study focused on three interrelated areas, including: (a) the use of copyright-protected works as training data for generative AI models; (b) the creation of new content by these systems and the legal challenges they present; and (c) liability frameworks for copyright infringement.[21]

39 The study concluded that there is a consensus that introducing a specific liability regime for copyright infringement in the context of generative AI is unnecessary at this stage, as existing general rules on infringement are deemed sufficient. This approach differs fundamentally from the safe harbour framework referenced in US law, as the EU does not employ equivalent safe harbour provisions under its copyright regime. Instead, EU copyright law permits generative AI training as text and data mining ("TDM") unless copyright holders have expressly reserved the use of their works through appropriate opt-out declarations.

40 The EU's liability framework for generative AI service providers operates within the existing tort and copyright law structure, emphasising principles of knowledge, control, and reasonable preventive measures. However, unlike the US approach, which relies on safe harbour immunities, the EU system focuses on balancing legitimate TDM activities with rights holders' interests through opt-out mechanisms and transparency requirements.[22] This approach reflects the EU's preference for adapting existing copyright exceptions and limitations rather than creating new immunity frameworks, resulting in a system that places greater emphasis on rights holder consent and platform responsibility for implementing effective opt-out systems.

41 Both the US and EU base their judgments on the standards of generative AI service providers' duty of care, drawing on existing tort and copyright laws, and considering their specific business models and technical capabilities. Considering that the courts in China also continue to base their decisions on the general principles of tort law and existing copyright law, it can be concluded that existing laws are generally capable of addressing the challenges posed by generative AI, and no

20 European Parliament, *Generative AI and Copyright: Training, Creation, Regulation* (PE 774.095, 2025) <https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf> (accessed 18 October 2025).

21 European Parliament, *Generative AI and Copyright: Training, Creation, Regulation* (PE 774.095, 2025) at p 108 <https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf> (accessed 18 October 2025).

22 Tim W Dornis & Sebastian Stober, "Generative AI Training and Copyright Law" (21 February 2025) <https://arxiv.org/abs/2502.15858> (accessed 3 August 2025).

significant adjustments to existing rules are required.[23] Accordingly, the key to determining the copyright infringement liability of generative AI service providers lies in how to adjust and apply existing rules. To what extent is training generative AI models on copyrighted materials considered permissible or infringing under Chinese copyright law? When a generative AI platform produces infringing content, how should the service provider's fault be defined to determine its liability? More specifically, under what circumstances should a breach of duty of care be deemed?

III. Copyright duty of care: training phase

A. Copyright limits and role of fair use in artificial intelligence training

42 In the case of generative AI, the training phase typically involves large-scale copying and storage of diverse content, including books, newspaper articles, songs, photographs and websites, into digital *corpora* used to build and refine the AI's capabilities.[24] Creating a training *corpus* through web scraping or database extraction typically requires reproducing and storing protected works, even if those works are no longer recognisable in the model's outputs.[25]

43 Therefore, at the core of the training-phase inquiry lies the question of whether copyrighted works may be used to train generative AI models without the authorisation of rights holders. This issue probes the boundaries of copyright's exclusive rights and its statutory limitations and exceptions. In the US, more than forty lawsuits have been filed against AI companies since 2022, in which rights holders alleged that AI companies have trained large language models using their works without consent or compensation. The outcomes of these cases are expected to have far-reaching implications for the AI industry.[26]

---

23 Wang Liming, "Principles of Attribution and Fault Determination in Generative Artificial Intelligence Tort" (2025) 4 *China Law Review* 15 at 16.

24 European Parliament, *Generative AI and Copyright: Training, Creation, Regulation* (PE 774.095, 2025) at pp 29–30 <https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf> (accessed 18 October 2025).

25 US Copyright Office, *Copyright and Artificial Intelligence – Part 3: Generative AI Training* (6 May 2025) at p 13.

26 Alex Reisner, "Judges Don't Know What AI's Book Piracy Means: Can AI Companies Keep Stealing Books to Train Their Models?", *The Atlantic* (14 July 2025) <https://www.theatlantic.com/technology/archive/2025/07/anthropic-meta-ai-rulings/683526/> (accessed 23 July 2025).

44 Under Chinese copyright law, authors enjoy broad rights over their original works, including but not limited to reproduction, distribution, dissemination through information networks, and adaptation. On its face, copying thousands of books or scraping millions of images to feed into an AI training dataset could implicate the reproduction right, and potentially the adaptation right if the training process is construed as producing latent derivative representations of the works. At the same time, China's copyright regime also recognises certain exceptions that allow the use of copyrighted materials without prior authorisation, functionally analogous to doctrines of fair use or fair dealing. Specifically, Art 24 of the Copyright Law of the People's Republic of China, as amended in 2020, sets out a closed list of 12 enumerated "fair use" scenarios, including private study, research, news reporting, and classroom teaching, among others, provided that such use does not conflict with the normal exploitation of the work and does not unreasonably prejudice the legitimate rights and interests of the copyright holder. Notably, Art 24(13) contains a catch-all provision for "other circumstances prescribed by laws and administrative regulations," which potentially leaves room for administrative rule-making or judicial interpretation to extend the scope of permissible uses.[27] In judicial practice, Chinese courts have repeatedly expanded the scope of fair use beyond the enumerated scenarios in Art 24, often invoking the four-factor test under US copyright law as a flexible standard to compensate for legislative gaps.[28]

45 Unlike the open-ended fair use doctrine in US copyright law, Chinese copyright law currently does not include AI training in its copyright exception clause. Indeed, TDM[29] or AI training is not covered by any of the 12 enumerated exceptions, and scholars note that whether TDM/AI training should qualify is a matter for the legislature, not for *ad hoc* judicial interpretation or creation. In other words, Chinese courts cannot simply declare a new fair use category unless it fits under an existing one or a new regulation authorises it. Consequently, at first glance, a literal reading of current law suggests that unlicensed AI training using copyrighted content is *prima facie* infringement in China.

---

27 In the 2014 draft amendment to the Copyright Law of the People's Republic of China, Art 43(13) provided for "other circumstances" and removed the restriction to those explicitly set by laws and administrative regulations, thereby leaving room for the construction of an open-ended fair use doctrine. Unfortunately, this provision was not ultimately adopted.

28 Zhang Chenguo, "Interpreting the 'Three-Step Test' and 'Fair Use'" (2016) *Global Law Review* 5 at 20.

29 The term "TDM" refers to automated analytical techniques used to extract patterns, trends, or correlations from large datasets, typically for scientific or informational purposes.

46 Nevertheless, Chinese authorities have taken a pragmatic stance, showing flexibility in practice. Instead of cracking down on AI companies for training data usage, the Interim Measures for the Management of Generative AI Services emphasise *ex post facto* regulation over pre-emptive restriction, focusing on governing AI's outputs and impacts while leaving the legal nature of the training phase somewhat "deliberately vague".[30] The rationale is straightforward: imposing rigid rules on training data too early could stifle AI development at its inception and hinder broader innovation. Thus, in practice, Chinese enforcement has remained lenient toward AI training itself, so long as it does not cause direct market harm.

47 The Hangzhou Internet Court's recognition of a *de facto* fair use exception for AI training in *Hangzhou Ultraman* provides an important reference point for addressing copyright implications of generative AI. In its reasoning, the court distinguished between four stages of the technology: data input, data training, content output, and content use, and applied a differentiated standard of review: relatively lenient for data input and training, but stricter for production and use. The court acknowledged that the development of generative AI inevitably involves ingesting vast amounts of training data, including third-party copyrighted works. According to the court, the purpose of using such works at the training stage is primarily to analyse underlying ideas, linguistic features, and stylistic elements to extract rules, structures, and patterns that facilitate subsequent transformative creation. Crucially, this process is not intended to replicate the original expressive aspects of the works. Instead, it generally involves only temporary, process-based copying, without any public disclosure of the underlying materials. Relying on principles such as non-expressive use, temporary or process-based copying, and the idea-expression dichotomy, the court held that, absent evidence that training is directed at reproducing protected expression, interfering with normal exploitation, or unreasonably harming the legitimate interests of rights holders, such use may qualify as fair use. Effectively, the court applied a harm-based test that echoes the three-step test[31] in international

30 In the draft version of the Interim Measures for the Management of Generative Artificial Intelligence Services, the provisions relating to intellectual property were notably stricter. For example, Art 4 required providers of generative AI products or services to comply with laws and regulations and refrain from infringing intellectual property rights, while Art 7 explicitly placed responsibility on providers to ensure that their pre-training and fine-tuning data did not contain IP-infringing content. However, these stringent requirements were subsequently weakened or removed in the final version of the measures.

31 The three-step test requires that exceptions: (a) apply only to certain special cases; (b) do not conflict with the normal exploitation of the work; and (c) do not unreasonably prejudice the legitimate interests of the rights holder.

copyright law. This reasoning represents a significant judicial innovation: it reflects a tolerant stance toward the use of copyrighted works in AI training and, in effect, sketches out a case law-based exception where statutory law remains silent.

48 Internationally, countries have adopted markedly different approaches to AI training, reflecting divergent priorities between protecting rights holders, fostering innovation, and ensuring legal certainty. The EU has taken a relatively restrictive stance on AI training. The EU Directive on Copyright in the Digital Single Market (Directive (EU) 2019/790)[32] introduced two TDM exceptions: Art 3 allows TDM for non-commercial research by public-interest institutions, and Art 4 extends the exception to commercial uses unless rights holders opt out through machine-readable means. However, because generative AI systems differ fundamentally from traditional TDM,[33] aiming not merely to detect correlations but to generate new expressive outputs, many scholars argue that large-scale AI training cannot be neatly classified as TDM or at least remains highly contested both technically and legally. The opt-out mechanism further complicates matters: it risks fragmenting training datasets while providing rights holders with little transparency or control. As a result, AI developers face legal uncertainty, while rights holders remain inadequately protected. By contrast, Japan has adopted a more permissive approach. The 2018 amendment to Japan's Copyright Act[34] introduced Art 30-4, a broad exception permitting the use of copyrighted works for purposes not involving the "enjoyment" of their expressive content. Under this "non-enjoyment" standard, computational uses such as machine learning or statistical analysis are not considered infringement so long as they do not reproduce the author's expression for human consumption. This doctrinal shift, treating copyright as infringed only when a work is used "as a work", reflects a policy orientation more supportive of AI innovation, granting wide latitude for data use in training. The US, meanwhile, has relied on its flexible fair use doctrine, which has historically played a central role in accommodating

32 Directive on Copyright and Related Rights in the Digital Single Market, Directive (EU) 2019/790, [2019] OJ L 130/92.

33 While TDM systems are designed to extract factual or semantic insights from data, generative AI models, particularly those based on transformer or diffusion architectures, are trained to synthesise new outputs by internalising expressive features of the input content. Through training, these models construct complex, multi-dimensional parameter spaces that capture style, structure, and composition, enabling them to produce outputs that may closely approximate original creative works. This transition from mere extraction to expressive recombination represents a fundamental departure from the analytical logic underlying the TDM exceptions, moving beyond the limited purpose of analysis toward the reproduction of expressive elements.

34 Copyright Act (Act No 48 of 1970) (as amended up to Act No 33 of 2023) (Japan).

technological innovation, from search engine indexing to the Google Books project. Recent case law suggests that US courts are extending this permissive attitude to AI training. In two rulings issued in mid-2025, one against Anthropic and another against Meta, federal judges held that using copyrighted books to train large language models constituted a "transformative" use, producing a fundamentally different product that does not directly compete with the original works.[35] Although the judges diverged in their reasoning, both concluded that the act of training itself falls within the ambit of fair use, signalling a pragmatic and relatively tolerant judicial approach toward AI model development.

49 To ensure legal certainty and strike an appropriate balance between AI innovation and copyright protection, China needs to clarify and codify copyright rules applicable to AI training. The key challenge is to preserve the socially beneficial aspects of AI training, including innovation, creativity, and new forms of expression, while preventing practices that genuinely undermine authors' rights, such as reconstructing or disclosing entire copyrighted works from a model. This paper argues that training generative AI models on copyrighted works should be recognised as a fair use-style exception under Chinese copyright law: a non-consumptive, non-expressive, and analytical use comparable to TDM. Such treatment should apply only if three conditions are satisfied: (a) the training does not substitute for the original works or impair their normal exploitation; (b) it relies on lawful access to the underlying materials; and (c) it embeds robust safeguards against memorisation and *verbatim* reproduction. In short, Chinese copyright law should regard AI training as conditional fair use, which is permitted only if developers fulfil their copyright alignment obligations.

50 Firstly, as a machine learning process, AI training typically engages copyrighted works in a non-expressive manner. The system extracts ideas, facts, and statistical patterns rather than reproducing protectable expression. OpenAI, for example, emphasises that its models are trained to discern patterns, linguistic structures, and contextual relationships, not to replicate works for public consumption.[36] This reflects the "non-expressive use" theory, widely discussed in international debates, which holds that AI training does not actually engage the

35 Alex Reisner, "Judges Don't Know What AI's Book Piracy Means: Can AI Companies Keep Stealing Books to Train Their Models?", *The Atlantic* (14 July 2025) <https://www.theatlantic.com/technology/archive/2025/07/anthropic-meta-ai-rulings/683526/> (accessed 23 July 2025).

36 OpenAI, "OpenAI's proposals for the US AI Action Plan" (13 March 2025) <https://openai.com/global-affairs/openai-proposals-for-the-us-ai-action-plan/> (accessed 25 July 2025).

expressive dimension of works and therefore may fall outside the scope of copyright.

51 Secondly, by analogy to human learning, AI training also resonates with the doctrine of transformative fair use. Training an AI model resembles a person reading extensively to inform the creation of new works: the process is creative and transformative, not a market substitute. This understanding was implicitly reflected in the Hangzhou Internet Court's fair use decision, which reaffirmed the idea-expression dichotomy. The court recognised that extracting styles or patterns (ideas) is permissible, while direct reproduction of artworks (expression) is not. The court even rejected theories like "retained expression" or "retained style" that would claim a model unlawfully appropriates an author's style, reasoning that such theories risk confusing unprotectable ideas with protected expression.

52 Thirdly, however, the legality of AI training must be distinguished from the legality of obtaining training data. Even if training itself is characterised as fair use or non-infringing, acquiring works through unauthorised channels constitutes a separate infringement of the reproduction right. This distinction is critical, as lawful access is a threshold requirement for any fair use claim. Developers must demonstrate that their datasets are built on legitimate sources, through licenses, public domain materials, or statutory exceptions, and that any copying is justified under copyright's permitted uses. In practice, this separation underscores a two-step compliance framework, with the first aspect being the lawful acquisition of training data, and the second being the lawful use of that data in training.

53 In summary, there is growing recognition in China that AI training may constitute a legitimate form of fair use of copyrighted materials. However, the absence of a clear statutory exception forces AI developers to navigate a precarious patchwork of implied tolerance, inconsistent judicial decisions, and *ad hoc* risk-management strategies, such as relying on licensed or public-domain datasets. While this fragile status quo may suffice in the short term, it is ultimately unsustainable: legal uncertainty chills investment, slows innovation, and undermines China's strategic ambition to lead in AI. If the law continues to lag behind technological practice, AI developers will be compelled to operate under constant legal risk, while rights holders remain without a coherent framework for protection. To resolve this impasse, China must move beyond piecemeal approaches and explicitly address the issue by enacting a tailored statutory exception for AI training in future revisions of its copyright law, one that provides legal certainty while safeguarding the legitimate interests of authors.

B. Copyright alignment obligations: scope and practical implications

54 The recognition that training generative AI systems on copyrighted materials may qualify as fair use or a copyright exception does not imply a general exemption from copyright norms for AI developers and providers. On the contrary, this paper argues that the legality of AI training must be conditioned on a copyright alignment obligation: the responsibility of AI developers to proactively design training processes that embed copyright compliance into the architectures of their AI models. This means building in safeguards during and immediately after training to reduce the risk that AI models will produce infringing materials when deployed. Furthermore, this obligation bridges the gap between legal permissibility and practical responsibility, ensuring that AI innovation does not undermine the legitimate interests of rights holders. In essence, while the training process itself may be excused, the outcome of training must be managed so that the model does not become a tool for piracy.

55 Technically, this copyright alignment obligation corresponds with the broader notion and practice of value alignment in the AI industry. As one of the central approaches to technical AI governance,[37] value alignment complements and reinforces legal and ethical governance of AI. Indeed, future AI governance may rely increasingly on technical mechanisms rather than traditional legal and regulatory frameworks alone. Accordingly, value alignment has become a cornerstone of responsible AI development worldwide, seeking to ensure that the purposes and behaviours of AI models conform to human values, ethical norms, and safety requirements. Developers already employ alignment methods such as reinforcement learning from human feedback and principle-based approaches like constitutional AI[38] to mitigate harmful, biased, dishonest, or misleading outputs. For Open AI, one objective of its AI model's alignment training during the post-training process is specifically to prevent content regurgitation, alongside other goals such as reducing harmful content, hallucinations, bias, and emergent risks.[39]

---

37 In general, technical AI governance refers to the application of technical expertise, methods, and tools to support, implement, and enhance the effective and reliable governance of AI systems in practice. It complements legal, policy, and ethical approaches by providing concrete technical mechanisms and analytical frameworks that help translate governance principles into enforceable and operational safeguards.

38 Yuntao Bai *et al*, "Constitutional AI: Harmlessness From AI Feedback", *Anthropic PBC* (15 December 2022) <https://www.anthropic.com/research/constitutional-ai-harmlessness-from-ai-feedback> (accessed 1 August 2025).

39 Aaron Jaech *et al*, "OpenAI o1 System Card" (21 December 2024) <https://arxiv.org/abs/2412.16720> (accessed 1 August 2025).

56 The EU has moved further toward codifying training-phase copyright duties through the Artificial Intelligence Act[40] and its accompanying General-Purpose AI Code of Practice[41] ("GPAI Code"). Article 53 of the Act requires general-purpose AI providers to implement a copyright compliance policy, maintain technical documentation for regulators, make relevant information available to downstream deployers, and publish a sufficiently detailed summary of their training content. These obligations effectively hardwire copyright alignment into the training stage itself. The Commission-endorsed GPAI Code translates these legal duties into operational guidance: it instructs providers to ensure lawful data sourcing, including respect for TDM opt-outs and avoidance of infringing repositories, to disclose data provenance, and to establish complaint and remedy mechanisms for rights holders. Of particular significance is the mandatory template for training-content summaries, which requires providers to specify data modalities, enumerate source types, such as public, private, scraped, and synthetic, and, for scraped data, provide narrative descriptions including the most relevant domains used. By institutionalising transparency and lawful sourcing requirements, the EU framework elevates copyright alignment from an aspirational best practice to a regulatory expectation, thereby anchoring compliance directly in the design and governance of frontier AI models.

57 Copyright alignment can be seen as a natural extension of this paradigm, functioning as a technical form of copyright governance for generative AI. In the training and especially post-training phases, this involves at least three elements: (a) structuring learning processes to minimise memorisation and *verbatim* reproduction of copyrighted materials; (b) fine-tuning models with copyright-specific guardrails to reduce infringement risks *ex ante*; and (c) calibrating refusal mechanisms and output filters to avoid generating content substantially like protected works.

58 Properly conceived, copyright alignment during training constitutes a legally significant form of AI value alignment. It transforms the principle of fair use into a defensible compliance framework, mitigates the risk of downstream infringement through memorisation, and provides auditable evidence of good-faith governance. In this way, copyright alignment reconciles the need to safeguard innovation with the imperative to protect authors' rights, anchoring the future of generative

40 Regulation (EU) 2024/1689, [2024] OJ L 2024/1980.

41 European Commission, "General-Purpose AI Code of Practice Now Available", press release (10 July 2025) <https://digital-strategy.ec.europa.eu/en/news/general-purpose-ai-code-practice-now-available> (accessed 30 October 2025).

AI in a framework of conditional legality in respect of using copyrighted materials for AI training. Furthermore, the alignment obligation essentially bridges the training and deployment phases: it is undertaken during training (or immediately after) but with an eye toward compliance in deployment. By instilling respect for copyright into the model's very behaviour, providers can better navigate the thin line between learning from existing works and unlawfully exploiting them.

59 This copyright alignment obligation is both reasonable and technologically grounded, because empirical studies have confirmed that large language models are capable of memorising and reproducing *verbatim* passages from training *corpora*, including copyrighted works and personally identifiable information.[42] This evidence establishes a duty to apply state-of-the-art anti-memorisation techniques at the training stage, such as dataset deduplication, contamination audits, insertion of "canaries" to test for memorisation, regularisation and early stopping, rare-sequence monitoring, and refusal rules calibrated through post-training alignment. By integrating such measures into the alignment stack, developers can demonstrate that copyright is not an afterthought but a design principle.

60 At the same time, copyright alignment must be a reasonable and proportionate obligation. Its scope should evolve in tandem with technological advancements and industry practices, rather than imposing an unrealistic standard of perfection. A prominent example is the problem of adversarial attacks. Even if developers adopt robust copyright alignment measures, malicious end users may still jailbreak or manipulate models to produce infringing outputs. In such cases, providers should not be deemed to have failed their copyright alignment obligations so long as they have taken reasonable steps consistent with best practices. This reflects the dynamic nature of the obligation: it requires diligence and continuous improvement, not absolute prevention.

61 Normatively, copyright alignment duties can also function as a liability shield. If a provider can demonstrate that it adopted reasonable preventive measures and integrated state-of-the-art safeguards, this strengthens its fair use defence in court and may even avert liability. Chinese case law already hints at this trajectory: courts have examined whether providers fulfilled their duty of care when assessing fault and damages. In effect, adherence to copyright alignment obligations could

42 A Feder Cooper *et al*, "Extracting Memorized Pieces of (Copyrighted) Books From Open-Weight Language Models" (18 May 2025) <https://arxiv.org/abs/2505.12546?> (accessed 1 August 2025).

become akin to a form of safe harbour – not absolute immunity, but a means to limit exposure by demonstrating good-faith efforts to avoid infringement. In this sense, copyright alignment is not a substitute for fair-use arguments but their operational foundation. Without demonstrable alignment, a fair-use defence remains fragile; with it, the defence becomes credible.

62 Taken together, copyright alignment obligations represent both a legal and technical innovation. They allow courts and regulators to tolerate AI training as fair use while ensuring that the resulting models are not misused as engines of infringement or piracy. They also give providers a clear compliance framework, balancing innovation with the protection of rights holders. Ultimately, copyright alignment operationalises the principle of "conditional legality": AI training on copyrighted materials may be permissible, but only if developers embed respect for copyright directly into their systems through alignment safeguards, for instance, copyright protection safeguards. This conditionality ensures that the growth of generative AI remains consistent with the rule of law and with broader societal values.

IV. Copyright duty of care: deployment phase

A. Legal status of generative artificial intelligence service providers

63 The deployment phase of generative AI platforms has generated significant legal controversy regarding their classification as service providers. This classification bears critical implications for liability frameworks, particularly in copyright infringement cases. The central question revolves around whether generative AI service providers should be categorised as internet content providers – who have breached their duty of care and should face primary infringement liability – or internet service providers in the narrower sense,[43] subject to contributory

43 In Chinese law, both the Regulations on the Protection of Information Network Dissemination Rights (信息网络传播权保护条例) (effective 1 July 2006, revised 30 January 2013) (special law) and the Civil Code of the People's Republic of China (中华人民共和国民法典) (effective 1 January 2021) (general law) treat internet platforms that directly provide content as "internet service providers". In other words, the distinction in Chinese law between platforms that provide content and platforms that provide services is also made within the broader framework of internet service providers, with individual determinations based on the specific actions of the platforms. This differs from the terms used in the US Digital Millenium Copyright Act 17 USC (US). Therefore, in this case, following Chinese legislative terms, all platforms that rely on the internet to provide internet services are referred to as internet service providers in a broader sense, while platforms that

(cont'd on the next page)

infringement standards when they violate their duty of care. Therefore, determining the legal status of generative AI service providers lies in analysing their role in the deployment phase of content generation, specifically whether they provide content or technical support services. However, given the technical and operational characteristics of the generative AI platform, this distinction is difficult to make.

64 Generative AI platforms exhibit hybrid characteristics that complicate traditional classification schemes across three dimensions. The primary aspect is that generative AI platforms draw from dual information streams in terms of information sources. They utilise vast pre-training datasets sourced from diverse origins, including internet content, which parallels the information foundation of traditional content providers. Simultaneously, they incorporate user-generated prompts and uploaded content that serve as crucial determinants in output generation. This dynamic, user-responsive content creation, conversely, distinguishes generative AI platforms from traditional content providers who rely primarily on pre-produced, curated material. Further, generative AI platforms participate in content creation through algorithmic processing rather than active editorial intervention characteristic of traditional content providers. While this involvement cannot be dismissed as purely passive, it fundamentally differs from the deliberate content curation and editorial control exercised by conventional internet content providers. With respect to output delivery, therefore, unlike traditional content providers who deliver pre-existing content, generative AI platforms generate novel content through interactive user engagement. This dynamic generation process creates content that is neither entirely pre-produced by the service provider nor solely user-created.

65 The legal status of generative AI service providers has sparked controversy due to technological and operational differences between them and traditional internet content providers, as well as internet service providers in the narrower sense. Proponents of content provider classification argue that generative AI platforms possess sophisticated content creation capabilities that leverage extensive language model knowledge through fine-tuned instructions.[44] Under this view, supported by the "digital gatekeeper theory" and "security obligation theory," generative AI providers should bear content-review responsibilities

provide content are referred to as internet content providers, and platforms that only provide technical services subject to safe harbour principles are referred to as internet service providers in a narrower sense.

44 Gao Yang, "Justification of the Duty of Care for Content Review by General Artificial Intelligence Providers" (2024) 1 *Oriental Law* 189 at 191.

equivalent to traditional content providers.[45] This approach is exemplified in *Guangzhou Ultraman*, where the court determined that the generative AI platform's access to copyrighted works, combined with substantial similarity between generated content and protected works, constituted primary infringement. Furthermore, alternative scholarly perspectives contend that generative AI providers primarily function as sophisticated information-retrieval tools, storing and processing user-directed information within their systems. Under this framework, generative AI providers should be classified as internet service providers in the narrower sense and should be subject to established safe harbour provisions (*ie*, "notice and take down") and red flag standards rather than proactive content-review obligations.[46]

66 While the process of content generation in the deployment phase possesses specific characteristics of content provision, generative AI providers should not be considered internet content providers. Reflecting on the legislative purpose and the interplay between legal and technological development, the DMCA, which first established the safe harbour rule, separated internet service providers from internet content providers. The key reason is to strike a balance between the interests of copyright holders and technological development.[47] Specifically, imposing content review as a duty of care on non-content service providers would require them to invest significant human and material resources in minimising the generation of infringing content – significantly increasing operating costs and even hindering the free flow of information. The costs and technological development barriers associated with requiring internet service providers to review content in the context of traditional internet service provision are already difficult to mitigate. This is compounded by the added challenges faced by generative AI service providers in generating content, given the exponential growth in the number of information sources and content provision. If generative AI service providers are required to assume the same responsibility as internet content providers, the resulting high operating costs would be detrimental to technological innovation, particularly for small and micro-sized innovative enterprises.[48] Thus, generative AI service providers should not be classified as content service providers and should not be held liable based on excessively high content review obligations. The draft version of the Measures for the Administration of Generative AI

45 Gao Yang, "Justification of the Duty of Care for Content Review By General Artificial Intelligence Providers" (2024) 1 *Oriental Law* 189 at 195–196.

46 Xiong Qi, "Copyright Infringement Liability of Generative Artificial Intelligence Platforms" (2025) 2 *Global Law Review* 23 at 31.

47 Melville B Nimmer, *Nimmer on Copyright* (Matthew Bender Elite Products, 1999).

48 Liang Zhiwen, "The Characterisation of Generative AI Output and Its Liability Rules Under Copyright Law" (2025) 6 *Law Science* 129 at 139.

Services once stipulated that organisations and individuals (hereinafter referred to as "providers") offering chat, text, image, or audio generation services through generative AI products – including those providing programmable interfaces for others to generate such content – would bear the same responsibility as the producers of the generated content. However, this provision was ultimately removed from the final version of the Measures.

67 Furthermore, from the perspective of interactions between generative AI and users, although the generative AI service provider has technical control over model training, the generation of infringing content during the deployment phase is user-driven, with user input of prompt words triggering subsequent content generation. Without active user participation, infringing content would essentially not be generated. In other words, the generative AI service provider passively participates in content generation in the model deployment phase. Therefore, it should not be categorised as an internet content provider, and should not bear primary liability for infringing content.

68 The EU's *Generative AI and Copyright* report also emphasises this point. Under current EU law, generative AI service providers are generally not liable for users' infringing actions unless they facilitate or assist in the infringement.[49] Consequently, the EU currently limits the infringement liability of generative AI service providers to contributory infringement when they breach their duty of care.[50] Accordingly, based on the specific operational model of the generative AI platform, generative AI service providers should not be classified as internet content providers and should not bear primary infringement liability.

69 Due to the unique technical and operational characteristics of generative AI platforms, they should be recognised as a new category of internet service providers within the broader framework of intermediary service regulation. While generative AI service providers do not fit neatly into the traditional category of internet content providers, their services entail complex interactive processes through which users participate in the generation of content. In such interactions, the roles of content producer, user, and potential rights holder or infringing party may overlap or become indistinguishable, thereby complicating the

49 Digital Services Act, Regulation (EU) 2022/2065, [2022] OJ L 277/1.

50 European Parliament, *Generative AI and Copyright: Training, Creation, Regulation* (PE 774.095, 2025) at p 108 <https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf> (accessed 18 October 2025).

allocation of legal responsibility for generated outputs.[51] This makes it difficult to directly apply safe harbour and red flag rules in the way that it is traditionally applied with regard to internet service providers. In particular, the "necessary deletion measures" (or "notice and necessary measures" under the Chinese copyright regime) required under the safe harbour and red flag rules to prevent infringement are challenging to implement in the actual application of generative AI services.[52] Moreover, the information transmission capabilities of generative AI service providers are stronger than those of traditional network service providers, hence the traditional level of duty of care should not be directly applied to the former. Instead, the duty of care requirements for generative AI service providers should be appropriately tailored in accordance with its technical and operational characteristics.

B. "Necessary measures" under duty of care

70 The copyright duty of care for generative AI service providers is multifaceted, extending beyond passive hosting to active engagement with content creation. At the deployment stage, this duty is a proactive and reactive framework designed to mitigate infringement risks stemming from user-platform interaction.

(1) Systematic filtering

71 The deployment stage of generative AI systems necessitates a dual-layered approach to copyright protection through both prompt filtering and outcome screening mechanisms. This framework addresses the unique challenges posed by real-time content generation.

72 In this context, generative AI service providers bear a heightened duty of care to implement sophisticated prompt filtering systems that identify potentially infringing requests before content generation occurs. This heightened duty stems from the technological capability principle, which requires entities with advanced analytical capabilities to employ reasonable measures matching their technical level to prevent foreseeable harm. As a result prompt filtering has become a cornerstone

51 Wang Liming & Bao Dingyurui, "On the Analogical Application of the 'Notice' Rule to Torts Arising From Generative Artificial Intelligence Outputs" (2025) 4 *Journal of Comparative Law* 1 at 3.

52 Bi Wenxuan, "On the Duty of Care of Generative Artificial Intelligence Service Providers" (2025) 2 *Journal of Soochow University Philosophy & Social Science Edition* 90 at 93.

of responsible AI deployment.[53] Specifically, effective prompt filtering, built on multidimensional analysis, encompasses explicit copyright references, contextual inference of infringing intent, and pattern recognition of circumvention attempts.

73 These mechanisms must balance effective preventive measures with creative freedom, avoiding overly broad restrictions that would impede legitimate transformative uses. Legal frameworks should acknowledge that no filtering system can prevent all potential infringement, particularly in cases involving fair use considerations that require nuanced human judgment.[54]

74 Furthermore, complementing prompt filtering, output filtering provides essential secondary protection through comprehensive similarity detection before content is delivered. This mechanism addresses cases where potentially infringing content emerges despite non-problematic prompts, due to unexpected associations within training data or emergent algorithmic behaviours. The challenge lies in implementing output filtering that operates within acceptable response time constraints for real-time generation while maintaining high accuracy in detecting substantial similarity.[55]

75 In addition, this mechanism must integrate with regularly updated copyright databases and employ both hash-based matching for exact reproductions and semantic analysis for substantial similarity assessment. The duty of care requires reasonable efforts to identify clear infringement cases while providing appropriate safe harbour protection for good faith compliance efforts, even when perfect detection proves technologically unfeasible.

76 Generative AI providers' active participation in content generation creates enhanced duty of care obligations compared to traditional internet service providers. Recent cases such as *Tremblay v OpenAI Inc*[56] and *Silverman v OpenAI Inc*[57] demonstrate growing judicial scrutiny of AI systems' capacity to reproduce copyrighted

53 Markus Anderljung, Julian Hazell & Moritz von Knebel, "Protecting Society from AI Misuse: When Are Restrictions on Capabilities Warranted?" (2025) 40 *AI & Society* 3841.

54 Dan L Burk, "Algorithmic Fair Use" (2019) 86 *University of Chicago Law Review* 283.

55 Altynbek Amirzhanov, Cemil Turan & Alfira Makhmutova, "Plagiarism Types and Detection Methods: A Systematic Survey of Algorithms in Text Analysis" (2025) 7 *Frontiers of Computer Science* 1504725.

56 *Tremblay v OpenAI Inc* No 3:23-cv-03223 AMO (28 June 2023, ND Cal) (US).

57 *Silverman v OpenAI Inc* No 3:23-cv-03416 (7 July 2023, ND Cal) (US).

works, emphasising the necessity for comprehensive filtering and screening mechanisms.

77 However, implementation must navigate competing policy objectives. While established companies such as Google, Facebook, Amazon and OpenAI have access to large collections of text and image data, imposing excessive filtering obligations could create a legal problem for new entrants and potentially stifle innovation and competition. The legal framework must therefore calibrate duties based on provider capabilities and market position, ensuring proportionate obligations that preserve innovation incentives.

78 The development of prompt filtering and output screening duties represents a critical evolution in copyright protection for the AI era, necessitating a careful balance between proactive prevention and technological feasibility while preserving space for legitimate creative expression and continued innovation in generative AI systems.

(2) Content identification through watermarking and attribution

79 The watermarking obligation serves dual functions: (a) risk prevention for downstream users; and (b) traceability for rights enforcement.[58] This duty recognises that AI-generated content may incorporate expressions from copyrighted training data, creating secondary infringement risks for users who subsequently reproduce or distribute the generated content.[59]

80 Specifically, generative AI service providers must implement robust watermarking systems that: (a) clearly identify content as AI-generated through visible or embedded markers; (b) provide traceability information linking generated content to specific generation sessions; (c) include appropriate disclaimers about potential copyright influences in training data; and (d) maintain watermark integrity across common content manipulation and distribution channels.

81 Content identification through watermarking will not increase the burden on generative AI service providers.[60] In fact, it serves multiple

58 European Parliament, *Generative AI and Copyright: Training, Creation, Regulation* (PE 774.095,2025) at p 108 <https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf> (accessed 18 October 2025).

59 Liang Zhiwen, "The Characterisation of Generative AI Output and Its Liability Rules Under Copyright Law" (2025) 6 *Law Science* 129 at 141.

60 Lin Beizheng, "On the Duty of Care of Generative Artificial Intelligence Service Providers" (2024) 10 *Journal of Law Application* 148 at 161.

stakeholder interests by enhancing rights holders' ability to identify and address downstream infringement, notifying users of content origin and potential copyright risks, ensuring transparency to the public, and safeguarding compliance with regulatory requirements.

82 This obligation is increasingly formalised in legal and regulatory frameworks. Specifically, China's Cyberspace Administration has established pioneering requirements through the Measures for the Identification of AI-Generated Content[61] and the national standard Cybersecurity Technology – Labelling Method for Content Generated By AI.[62] These frameworks provide systematic approaches to content identification that can serve as models for implementation. From the perspective of judicial practice, *Guangzhou Ultraman* emphasised the importance of content watermarking in enabling targeted rights protection measures. The court's recognition of watermarking to facilitate more effective enforcement actions supports the legal foundation for this duty of care component.[63]

83 China's approach represents a pioneering effort to establish systematic requirements for content labelling and watermarking. Combined with the attitudes expressed in the EU research report, it suggests a global trend toward mandatory transparency requirements for AI-generated content, transforming best practices into legal obligations.

(3) Progressive user guidance and behavioural monitoring

84 Generative AI service providers have a duty of care to proactively guide user behaviour and prevent infringement. As providers of interactive online services, their responsibility is not limited to *post hoc* reactions but should encompass the entire user journey. First, before the generative AI platform is used, service providers should issue clear disclaimers that prohibit users from entering prompts highly correlated with copyrighted works, thereby discouraging the generation of infringing outputs from the outset. Building on this pre-use warning, platforms should implement real-time monitoring and warning systems. Upon detecting prompts or behaviours with a high probability of generating infringing content, the system should issue dynamic warnings. If a user persists in attempts to generate infringing content, technical measures

61 (人工智能生成合成内容标识办法) (effective 1 September 2025).

62 (网络安全技术人工智能生成合成内容标识方法) (effective 1 September 2025).

63 *Shanghai Character License Administrative Co Ltd v Guangzhou Nianguang Network Technology Co Ltd* (Guangzhou Internet Court) (2024) Yue 0192 Min Chu (Civil preliminary trial) No 113 ((2024) 粤0192民初113号) at pp 20–21.

such as content blocking or temporary suspension of services should be implemented to prevent further unlawful activity.

85 Regarding intervention strategies based on user behaviour patterns, a graduated response policy is essential for fairness and effectiveness. First-time infringers should receive a warning, particularly for keywords or content that may lead to infringement. In contrast, repeat infringers should face escalating consequences, including access restrictions and account termination, aligning with established practices in online content moderation.[64] This approach mirrors the strategies of some search engines, which append infringement notices or court rulings to suspicious search results to promote user caution and compliance.[65]

(4) Mechanism for "notice and necessary measures"

86 Based on the characterisation of a generative AI provider as an intermediary platform, if a network user commits an act of infringement through the use of generative AI, that user would bear civil liability in accordance with Chinese copyright law. The generative AI service provider, acting merely as an intermediary, would not, in principle, be held liable for the user's infringing act. However, if, upon receiving notice from the infringed party, the provider fails to take necessary and timely measures – given the preliminary evidence of infringement and the nature of its service – it would bear corresponding civil liability pursuant to Art 1195 of the Civil Code of the People's Republic of China.[66] Therefore, in the post-infringement phase, generative AI service providers have a core duty to establish an efficient and equitable "notice and action" or "notice and necessary measures" mechanism. This framework has been further refined in judicial practice and centres on two main components.

87 The first is the duty of timely and effective action. Providers must create accessible complaint channels, including 24/7 online portals for infringement notices, and offer clear guidelines for rights holders. Upon receiving valid notice, the provider should promptly take necessary measures, such as deleting, blocking, or preventing the regeneration of the allegedly infringing content or its derivative outputs. The response time and efficiency of this process should align with industry standards. In practice, the term "necessary measures" includes deleting infringing content, disconnecting links, blocking relevant generation prompts,

64 Feng Xiaoqing, "Determination of Copyright Infringement Liability of Generative Artificial Intelligence Service Providers" (2025) 1 *Research on Rule of Law* 46 at 50.

65 Eugene Volokh, "Shenanigans (Internet Takedown Edition)" (2021) 2 *Utah Law Review* 237 at 292–296.

66 (中华人民共和国民法典) (effective 1 January 2021).

taking reasonable optimisation measures to reduce the generation of similar infringing content by the model, and issuing warnings, restricting, or terminating services for the infringing user.

88 The second component is the duty of fairness to users. To protect user rights, a complete "counter-notice and review" procedure should be established, enabling users to appeal the decision and present arguments on the non-infringing nature of the content. The platform is then obligated to legally forward and co-ordinate the handling of the appeal, ensuring due process.

89 This mechanism represents an extension of the safe harbour principle into the domain of AI, and its institutional value is increasingly being demonstrated in judicial rulings. For example, in *Guangzhou Ultraman*, the court ruled that a service provider's duty goes beyond merely deleting the specific infringing content. The court emphasised that a provider must also implement technical measures to prevent the system from continuously generating substantially similar infringing outputs. This decision clarified that a platform may be found liable if it only performs a superficial deletion without fundamentally blocking the regeneration of infringing content. Similarly, the Hangzhou Internet Court's ruling has clarified that timely necessary measures should include subsequent protection at the model level, rather than just the disposal of existing content.

90 Ultimately, this post-infringement mechanism provides an effective pathway for rights holders to assert their rights while also establishing clear conditions for service providers' liability exemptions. By expanding the interpretation of necessary measures, courts are compelling platforms to fulfil a more proactive and profound copyright protection duty. This helps to achieve a balanced ecosystem where intellectual property rights are respected, and the development of the AI industry can continue to flourish.

91 This integrated duty of care framework for generative AI deployment stages balances copyright protection with technological innovation through progressive user guidance, comprehensive content identification, and risk-proportionate remediation obligations. By incorporating lessons from legal precedents, regulatory developments, and technical constraints, this framework provides a practical foundation for copyright compliance in the rapidly evolving generative AI landscape.

92 In summary, when an infringed party seeks to hold a generative AI service provider jointly and severally liable for an infringement committed by a network user, the determination of the provider's

liability should be made through a comprehensive assessment of multiple factors. These include:

(a) the type and operational model of the service and the extent to which such service may facilitate or induce infringement;

(b) whether the provider has duly labelled generated or synthetic content in accordance with applicable national regulations and technical standards;

(c) whether the provider has engaged – manually or algorithmically – in the recommendation, ranking, selection, editing, or modification of the generated content; and

(d) the technical feasibility of preventive measures as well as the reasonableness of the actions taken in response.

Based on a holistic evaluation of these elements, it may be determined whether the provider "knew or should have known" of the infringing act within the meaning of Art 1197 of China's Civil Code.

V. Concluding remarks

93 Existing copyright principles – designed for traditional internet services – are ill-suited to the distinct legal issues raised by generative AI systems that co-create content through complex algorithms. This article argues that China can reconcile robust copyright protection with frontier AI innovation by adopting a dual-phase duty-of-care architecture, *via*: (a) conditioning the lawfulness of training on *ex ante* "copyright alignment,"; and (b) calibrating proportionate provider obligations *ex post* at the point of deployment. Considered together, recent Chinese cases and comparative regimes indicate that the core task is not to invent a new liability system for generative AI, but to operationalise established copyright doctrines through clear, technically grounded duties aligned with real-world capabilities and risks – thereby systematically mitigating infringement across the entire AI lifecycle.

94 Firstly, at the training stage, the most sustainable path is conditional legality: recognising large-scale training on copyrighted materials as a limited, non-consumptive use – only if developers satisfy baseline duties of lawful access, dataset governance, and anti-memorisation controls. In practice, this means auditable provenance for data sources; separation of data acquisition from model training; robust de-duplication and contamination auditing; empirical testing for regurgitation (*eg*, canaries, rare-sequence probes); and post-training alignment targeted at suppressing *verbatim* reproduction and close-style emulation. This approach transforms fair use arguments

from abstract doctrine into a verifiable compliance program, reducing downstream harm while preserving the innovation dividend of broad training.

95 Secondly, at the deployment stage, providers should be subject to graduated, capability-sensitive duties that reflect their active role in interactive generation but avoid collapsing them into full content-provider liability. A workable baseline includes:

(a) input controls (prompt classification and friction for high-risk queries);

(b) output controls (near-duplicate and semantic similarity checks under latency budgets);

(c) content identification (persistent watermarks or provenance signals resilient to common transformations) to warn users and support traceability; and

(d) processes for "notice and necessary measures"[67] that go beyond one-off deletions to address repeated generation of substantially similar content.

These measures should be assessed against what is technically reasonable for the provider's scale and architecture, recognising that perfect *ex ante* prevention is infeasible in an adversarial environment.

96 Thirdly, China's legislation and judicial policy can convert this architecture into legal certainty through a three-pillar package:

(a) statutory training exceptions with conditions – lawful access, non-substitution, and state-of-the-art anti-memorisation – paired with transparency duties for high-capability models (*eg*, model cards and high-level training-data summaries that protect trade secrets while enablingoversight);

(b) risk-based deployment obligations tiered by provider capability, business model, user scale, and market impact, with safe harbour-like mitigation where providers document and maintain reasonable controls; and

67 For generative AI providers, the "notice and necessary measures" obligation should, at a minimum, encompass: (a) assistance in cutting off dissemination of infringing content; (b) blocking of relevant generation prompts; (c) reasonable model-optimisation steps to reduce the risk of similar infringing outputs; and (d) warnings, restrictions, or service termination for the infringing user.

(c) standards and institutions – authoritative technical specifications (dataset governance, watermark robustness, regression tests for memorisation), third-party audit pathways, and co-ordinated enforcement among China's digital and information governance authorities, as well as courts *via* judicial interpretations.

97 Fourthly, to align incentives, China should complement duties with market mechanisms such as: (a) liability insurance linked to demonstrable compliance; (b) optional collective licensing or levy schemes for specific repertoires where transaction costs are high; and (c) public-private datasets with clean provenance to lower entry barriers for small to medium-sized enterprises. Together, these instruments reduce litigation uncertainty, create predictable compliance gradients, and curb the risk of entrenching only the largest firms.

98 Fifthly, the framework must remain evolutionary. Duty-of-care benchmarks should be periodically updated to reflect empirical evidence about memorisation, new watermarking and provenance standards, jailbreak patterns, and model capabilities. Regulators can use sandboxes and time-bound pilots (*eg*, "duty-light" regimes for research-only deployments under strict containment) to validate requirements before nationwide roll-out, ensuring that obligations track what is technically and economically achievable.

99 Finally, the broader governance dividend of a copyright duty of care is strategic: it enables China to lead in responsible AI development without sacrificing competitiveness. By anchoring legality in verifiable engineering practices during training and insisting on proportionate, auditable safeguards at deployment, policymakers can protect authors, empower users, and give developers a stable rule-of-law foundation on which to build. The destination is not zero risk but managed risk – a regime where creative ecosystems and generative models can co-evolve under clear, practicable, and future-proof rules.
