Using copyrighted content to train generative AI can be deemed fair following Ultraman infringement dispute

Xiaoquan (Claus) Zhang, 30 April 2025, first published by IAM

With the extensive adoption of AI-powered tools, copyrighted content is increasingly being used to train large language models (LLMs). Whether this qualifies as fair use is a hotly debated topic around the world.

In February 2025 a ruling handed down by the Hangzhou Intermediate Court in December 2024 came to light. The court found that using copyrighted content to train generative AI could be deemed fair, provided that there is no evidence proving that such use intends to plagiarise the original expression of the copyrighted works or has impeded the use of the originals or unreasonably prejudiced the copyright owner's legitimate interests.

Case background

Tsuburaya Productions is the copyright holder of Ultraman, a Japanese anime character, while Shanghai Character License Administrative (SCLA) – the plaintiff – is the exclusive licensee of Ultraman fine art works.

Small Design is an AI platform operator that enables users to create LoRA models by fine tuning an AI model on a custom dataset. Users uploaded images of Ultraman to this platform to train original models, which were then made available to other users for generating varied Ultraman-style content.
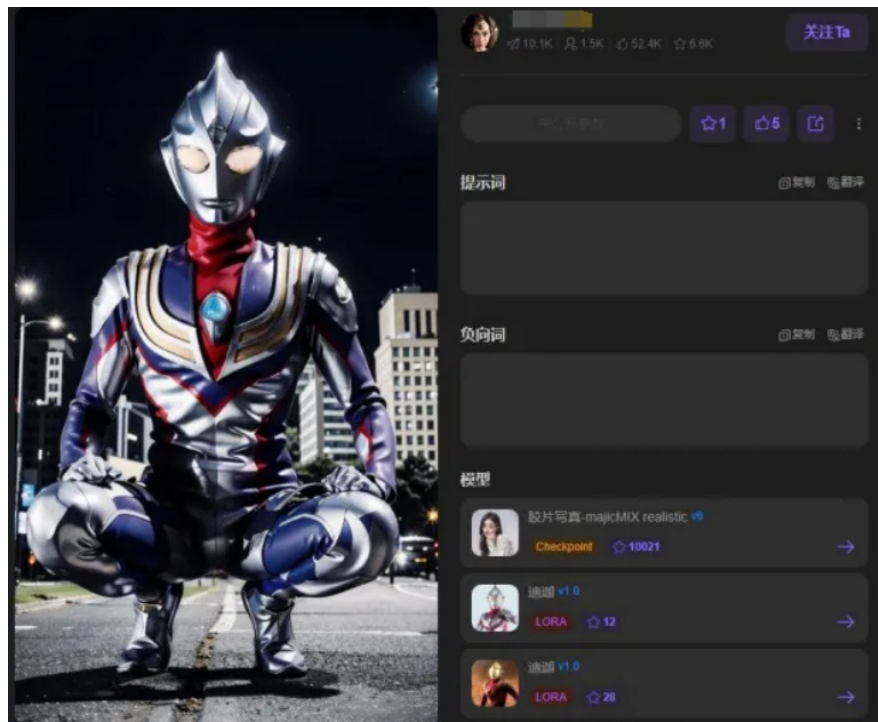
Figure 1. The Ultraman image generated by the accused platform

In 2024, SCLA sued Small Design on the grounds of copyright infringement and unfair competition. It requested cessation, deletion of all relevant material and data and damages (including reasonable costs) of 300,000 yuan.

The Hangzhou Internet Court dismissed the unfair competition claim as it found the defendant's business model and operations to be legitimate, and it held that the dispute would be governed by the Copyright Law rather than the Anti-Unfair Competition Law.

The court affirmed that the defendant, as a provider of generative-AI services, was not involved in direct copyright infringement. However, it held the defendant liable for contributory copyright infringement, ordering the generated Ultraman images to be deleted and cessation of the generation and publishing service pertaining to the images at issue. The court awarded damages and reasonable costs of 30,000 yuan, which is one tenth of the initial amount claimed by the plaintiff, and rejected the plaintiff's other requests.

The Hangzhou Intermediate Court upheld this decision on appeal.

Fair use of copyrighted content in AI training

The trial court used a two-pronged approach. Specifically, it proposed adopting a more lenient and inclusive assessment of data input and data-training actions of LLMs. However, it underlined the necessity of a rigorous assessment when it came to the output of LLM-generated content and its use. This bifurcated approach sharply contrasts Article 7(2) of the Interim Measures for the Management of Generative Artificial Intelligence Services, which stipulate that gen-AI service providers must

conduct pre-training, fine-tuning and other data-processing activities in accordance with the law – and must not infringe upon the legally protected rights of others – when IP rights are involved.

From the court's rationale, it can be inferred that the correlation between training and generation is not linear causality, as may be misconstrued by most people. The creation and evolution of generative AI requires the input of a massive amount of training data, which would inevitably include the copyrighted works of others. In principle, using these at the training stage is for learning, analysing and summarising prior works for the sake of transformative creation of new works later – without the intention to reproduce the originality of the copyrighted works. In general, data training merely temporarily retains the prior works in structural analysis of corpus data without making these available to the public during the training and generation processes.

The court therefore concluded that so long as the training process does not intend to reproduce the original works, interfere with their normal use or cause unreasonable harm to the copyright holder's legitimate interests, it may fall within the scope of fair use.

Key takeaways

Notably, Article 24 of the 2020 Chinese Copyright Law outlines the circumstances that constitute fair use of copyrighted works. Although the legislation does not explicitly address the use of copyrighted works for AI training, clause 13 (other circumstances provided by laws and administrative regulations) leaves the door open, should other laws or administrative regulations sanction fair use in the context of AI training in the future.

It is interesting that the Hangzhou Intermediate Court distinguished the platform's duty to dissuade user infringement during the model-training phase (input) by using legally sourced data and models from its obligations to prevent infringement at the content-generation stage (output) by utilising necessary mechanisms that are consistent with the level of technology at the time of the infringement. This distinction reflects a growing judicial understanding in China that a nuanced approach may be warranted in determining the legal liability of platform operators hosting AI-powered tools in the different phases of AI training and utilisation.

Training AI requires huge and diverse datasets, which often include copyrighted text, images and audio. Whether such use constitutes infringement remains a legal grey area worldwide. This case, which is China's first judicial recognition in this area, seems to insinuate a more tolerant attitude toward copyrighted inputs in AI training – but a stricter stance on infringing outputs. The decision could help to shape both legal interpretation and industry practices in China and beyond.