



AI Art: Myth vs Reality, What the \$1.5B Anthropic Deal and the Getty Ruling Actually Decided

\$1.5 billion did not buy Anthropic a license to train on books. It bought peace over 7 million pirated books sitting in a library, after the judge had already said the training itself might be fair use.

THE SHORT VERSION

Four 2025 decisions point the same way: how you got the data decides your liability, not what the model learned from it. Judge Alsup ruled in June 2025 that training Claude on the authors' works could qualify as fair use. Anthropic still paid \$1.5 billion, approved September 25, 2025, for retaining over 7 million pirated books. The UK High Court dismissed Getty's secondary copyright claim on November 4, 2025 because model weights store statistics, not images.



If you ship generative visuals, your legal exposure lives in procurement records and storage rather than in the model card. A pipeline built on licensed or lawfully acquired inputs survives both the Anthropic and the Getty reasoning. A pipeline built on a scraped archive someone kept “for reproducibility” does not, and no amount of fine-tuning hygiene fixes that afterwards. Believing otherwise costs real money: Anthropic’s bill came to roughly \$3,000 per book across about 500,000 downloaded books.

Getty did not settle whether training on scraped images is legal

THE CLAIM

Getty lost, so training on scraped images is settled law and the copyright fight is over.

THE REALITY

Getty’s broader copyright and database-right claims were withdrawn or dismissed before judgment, so the court never ruled on training. The [English High Court rejected only the secondary infringement claim](#) ([2025] EWHC 2863 (Ch), November 4, 2025), on the narrow basis that Stable Diffusion’s weights encode learned statistical relationships rather than storing Getty images. Stability was granted permission to appeal that very point on December 17, 2025.

Two details from that judgment matter more than the headline. First, the court accepted that an AI model can in principle be an intangible “article” under the UK Copyright, Designs and Patents Act 1988. The weights were not infringing copies, but the category door is open. Second, Getty did win something: a narrow trademark finding tied to Getty and iStock watermarks appearing in outputs. That is not broad liability for Stable Diffusion’s outputs, and it is not nothing either if your product surfaces recognisable marks.

Anthropic paid for retention, not for training



THE CLAIM

The Anthropic settlement proves training on copyrighted work is illegal and will cost everyone billions.

THE REALITY

Judge William Alsup ruled in June 2025 that training Claude on the authors' works could qualify as fair use. Liability came from retaining more than 7 million pirated books in a central library. The [\\$1.5 billion settlement approved on September 25, 2025](#) works out to roughly \$3,000 per book across about 500,000 downloaded books.

Ross lost on fair use even though users never saw the copied headnotes

The third myth is that invisible outputs protect you. In [Thomson Reuters v. Ross Intelligence](#) (February 11, 2025, Judge Stephanos Bibas, D. Del.), the amount factor favoured the defendant because users never saw the 2,243 copied Westlaw headnotes. Fair use was still rejected. Commercial use that substitutes for the original market decided the balance.

Ross is the one I would hand to an engineering team first, because it breaks the intuition most pipelines are built on. Invisible copying is still copying when the product you build competes with the thing you copied from. For an image pipeline the analogue is uncomfortable: a model trained to produce stock-style imagery, competing with the stock library it ingested, sits closer to Ross than to Anthropic's fair-use finding on training.

Can you copyright an image your model generated?

Partly, and only the parts a human actually authored. The US Copyright Office released Part 2 of its AI report, [Copyrightability](#), on January 29, 2025, and reached three conclusions worth memorising. Existing law is sufficient, so no new AI statute is coming from that direction. AI assistance does not disqualify a work; copyright protects the human-authored expressive elements, assessed element by element rather than across the whole file. And prompts alone are generally not enough for authorship, while human selection, arrangement, modification, or incorporation into a larger human-authored work can qualify.



Read that element-by-element rule carefully. A generated asset is not a single indivisible thing. The composition you arranged, the retouching you did, the layout you placed it into are yours; the raw generation underneath may be unprotectable. If your business model assumes exclusive rights over a library of prompt-only outputs, the Office has told you plainly that assumption is wrong.

✓ **The fix**

Keep an acquisition log before you keep a training log. For every corpus: where it came from, under what licence, when it was obtained, and what was deleted afterwards. Anthropic's exposure came from retention of the pirated library, so deletion policy is a liability control. Separately, record the human contribution to each shipped asset, because that record is what the Copyright Office's element-by-element test will be applied to.

Synthetic data only collapses a model when it replaces the real data

The remaining myth about AI imagery is technical rather than legal: that synthetic data poisons models irreversibly. Shumailov et al., *AI models collapse when trained on recursively generated data*, Nature vol. 631, pp. 755 to 759, July 24, 2024, is the paper everyone cites. The paper people do not cite is [Gerstgrasser et al., arXiv:2404.01413](#) (April 1, 2024), which found collapse is driven by *replacing* real data with synthetic. Accumulate synthetic data alongside retained real data and test error stays bounded.

That is a design instruction rather than a reassurance. Replacement kills you; accumulation does not. So your data retention policy is doing double duty, answering the legal question and the quality question at once, and the two pull in slightly different directions. You need to keep real data to avoid collapse, and you need provable rights to the real data you keep.

WHERE I LAND

My take: these myths persist because both sides of the debate benefit from them. "Training is legal" is a cleaner story for vendors than "acquisition is the liability", and "AI companies pay billions" is a cleaner story for rights holders than "a judge said the training was probably fine". Reading the four decisions together, the fight moves from model architecture to provenance paperwork, and the companies that lose will be the ones who cannot prove how a dataset arrived. Unconfirmed, and this is my prediction rather than anything a court has said: the

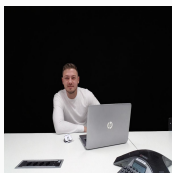


AI Art: Myth vs Reality, What the \$1.5B Anthropic Deal and the Getty Ruling Actually Decided

*Stability appeal on secondary infringement, granted December 17, 2025, is the one to watch, because the High Court already conceded that a model can be an “article” under the 1988 Act. If an appeal court takes the next step on what counts as an infringing copy, the **weights-are-just-statistics defence** gets considerably thinner.*

The commercial picture does not follow the legal one neatly either. I have written before about how [artists are winning in court and losing in the market](#), and the gap between those two outcomes is the thing most procurement conversations ignore. Winning a narrow trademark point, as Getty did, does not restore a licensing market.

So: audit where your training and reference images came from, write down what you deleted, and stop assuming a fair use finding on training covers how you obtained the files. If you have a pipeline where the provenance answer is genuinely unclear, I am [happy to talk it through](#) before your lawyers have to.



WRITTEN BY

Artur Markus

Software engineer in Basel building AI infrastructure and agentic systems, with a background in pharmaceutical automation, GxP compliance, and validation. I help teams turn AI from a chatbot into a reliable operational layer.

Book a meeting →