

Anthropic gives court authority to intervene if chatbot spits out song lyrics

Anthropic did not immediately respond to Ars' request for comment on how guardrails currently work to prevent the alleged jailbreaks, but publishers appear satisfied by current guardrails in accepting the deal.

Whether AI training on lyrics is infringing remains unsettled

Now, the matter of whether Anthropic has strong enough guardrails to block allegedly harmful outputs is settled, Lee wrote, allowing the court to focus on arguments regarding "publishers' request in their Motion for Preliminary Injunction that Anthropic refrain from using unauthorized copies of Publishers' lyrics to train future AI models."

Anthropic said in its motion opposing the preliminary injunction that relief should be denied.

"Whether generative AI companies can permissibly use copyrighted content to train LLMs without licenses," Anthropic's court filing said, "is currently being litigated in roughly two dozen copyright infringement cases around the country, none of which has sought to resolve the issue in the truncated posture of a preliminary injunction motion. It speaks volumes that no other plaintiff—including the parent company record label of one of the Plaintiffs in this case—has sought preliminary injunctive relief from this conduct."

In a statement, Anthropic's spokesperson told Ars that "Claude isn't designed to be used for copyright infringement, and we have numerous processes in place designed to prevent such infringement."

"Our decision to enter into this stipulation is consistent with those priorities," Anthropic said. "We continue to look forward to showing that, consistent with existing copyright law, using potentially copyrighted material in the training of generative AI models is a quintessential fair use."

This suit will likely take months to fully resolve, as the question of whether AI training is a fair use of copyrighted works is complex and remains hotly disputed in court. For Anthropic, the stakes could be high, with a loss potentially triggering more than \$75 million in fines, as well as an order possibly forcing Anthropic to reveal and destroy all the copyrighted works in its training data.

<https://arstechnica.com/tech-policy/2025/01/anthropic-gives-court-authority-to-intervene-if-chatbot-spits-out-song-lyrics/>