



A BAD SEED?

“Torrenting from a corporate laptop doesn’t feel right”: Meta emails unsealed

Meta's alleged torrenting and seeding of pirated books complicates copyright case.

ASHLEY BELANGER — FEB 6, 2025 4:26 PM |

178

Credit: Devonyu | iStock / Getty Images Plus



TEXT SETTINGS

Newly unsealed emails allegedly provide the "most damning evidence" yet against Meta in a copyright case raised by book authors alleging that Meta illegally trained its AI models on pirated books.

Last month, Meta admitted to torrenting a controversial large dataset known as LibGen, which includes tens of millions of pirated books. But details around the torrenting were murky until yesterday, when Meta's unredacted emails were made public for the first time. The new evidence showed that Meta torrented "at least 81.7 terabytes of data across multiple shadow libraries through the site Anna’s Archive, including at least 35.7 terabytes of data from Z-Library and LibGen," the authors' court filing said. And "Meta also previously torrented 80.6 terabytes of data from LibGen."

"The magnitude of Meta’s unlawful torrenting scheme is astonishing," the authors' filing alleged, insisting that "vastly smaller acts of data piracy—just .008 percent of the amount of copyrighted works Meta pirated—have resulted in Judges referring the conduct to the US Attorneys’ office for criminal investigation."

Seeding expands authors’ distribution theory

Book authors had been pressing Meta for more information on the torrenting because of the obvious copyright concern over Meta seeding, and thus seemingly distributing, the pirated books in the dispute.

But Meta resisted those discovery attempts after an order denied authors' request to review Meta's torrenting and seeding data. That didn't stop authors from gathering evidence anyway, including a key document that starts with at least one staffer appearing to uncomfortably joke about the possible legal risks, eventually growing more serious about raising his concerns.

"Torrenting from a corporate laptop doesn’t feel right," Nikolay Bashlykov, a Meta research engineer, wrote in an April 2023 message, adding a smiley emoji. In the same message, he expressed "concern about using Meta IP addresses 'to load through torrents pirate content.'"

By September 2023, Bashlykov had seemingly dropped the emojis, consulting the legal team directly and emphasizing in an email that "using torrents would entail ‘seeding’ the files—i.e., sharing the content outside, this could be legally not OK."

Emails discussing torrenting prove that Meta knew it was "illegal," authors alleged. And Bashlykov's warnings seemingly landed on deaf ears, with authors alleging that evidence showed Meta chose to instead hide its torrenting as best it could while downloading and seeding terabytes of data from multiple shadow libraries as recently as April 2024.

Meta allegedly concealed seeding

Supposedly, Meta tried to conceal the seeding by not using Facebook servers while downloading the dataset to "avoid" the "risk" of anyone "tracing back the seeder/downloader" from Facebook servers, an internal message from Meta researcher Frank Zhang said, while describing the work as being in "stealth mode." Meta also allegedly modified settings "so that the smallest amount of seeding possible could occur," a Meta executive in charge of project management, Michael Clark, said in a deposition.

Now that new information has come to light, authors claim that Meta staff involved in the decision to torrent LibGen must be deposed again because the new facts allegedly "contradict prior deposition testimony."

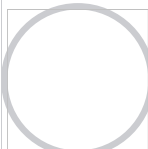
Mark Zuckerberg, for example, claimed to have no involvement in decisions to use LibGen to train AI models. But unredacted messages show the "decision to use LibGen occurred" after "a prior escalation to MZ," authors alleged.

Meta did not immediately respond to Ars' request for comment and has maintained throughout the litigation that AI training on LibGen was "fair use."


However, Meta has previously addressed its torrenting in a motion to dismiss filed last month, telling the court that "plaintiffs do not plead a single instance in which any part of any book was, in fact, downloaded by a third party from Meta via torrent, much less that Plaintiffs' books were somehow distributed by Meta."

While Meta may be confident in its legal strategy despite the new torrenting wrinkle, the social media company has seemingly complicated its case by allowing authors to expand the distribution theory that's key to winning a direct copyright infringement claim beyond just claiming that Meta's AI outputs unlawfully distributed their works.

As limited discovery on Meta's seeding now proceeds, Meta is not fighting the seeding aspect of the direct copyright infringement claim at this time, telling the court that it plans to "set... the record straight and debunk... this meritless allegation on summary judgment."



ASHLEY BELANGER
SENIOR POLICY REPORTER
Ashley is a senior policy reporter for Ars Technica, dedicated to tracking social impacts of emerging policies and new technologies. She is a Chicago-based journalist with 20 years of experience.



178 COMMENTS





MOST READ



- 1. “Torrenting from a corporate laptop doesn’t feel right”: Meta emails unsealed**
- 2. Boeing has informed its employees that NASA may cancel SLS contracts**
- 3. Stem cells used to partially repair damaged hearts**
- 4. NASA will swap Dragon spacecraft on the ground to return Butch and Suni sooner**
- 5. White House budget proposal could shatter the National Science Foundation**



Ars Technica has been separating the signal from the noise for over 25 years. With our unique combination of technical savvy and wide-ranging interest in the technological arts and sciences, Ars is the trusted source in a sea of information. After all, you don’t need to know everything, only what’s important.

MORE FROM ARS

[ABOUT US](#)
[STAFF DIRECTORY](#)
[NEWSLETTERS](#)

ARS VIDEOS
GENERAL FAQ
RSS FEEDS

CONTACT

CONTACT US
ADVERTISE WITH US
REPRINTS



DO NOT SELL MY PERSONAL INFORMATION

© 2025 Condé Nast. All rights reserved. Use of and/or registration on any portion of this site constitutes acceptance of our User Agreement and Privacy Policy and Cookie Statement and Ars Technica Addendum and Your California Privacy Rights. Ars Technica may earn compensation on sales from links on this site. Read our affiliate link policy. The material on this site may not be reproduced, distributed, transmitted, cached or otherwise used, except with the prior written permission of Condé Nast. Ad Choices