# Meta faces legal storm over pirated books used in AI training



The controversy surrounding artificial intelligence and its training data has escalated dramatically with revelations about Meta's practices. Recent investigations unveiled that the tech giant, helmed by Mark Zuckerberg, engaged in the controversial use of pirated literature from the Library Genesis database—commonly referred to as LibGen—serving as a significant source for its AI systems, including the Llama models. This incident has sparked outrage among authors and stakeholders in the literary community, leading to legal actions and passionately voiced critiques about the ethics of AI training methodologies.

The fallout from this discovery has been swift and severe. In January 2025, notable authors, including Ta-Nehisi Coates and Sarah Silverman, initiated a lawsuit against Meta, accusing the company of using their works without permission. Internal documents have suggested that Meta was aware of the pirated nature of the content, raising critical questions regarding not only the legality but also the ethical implications of using such datasets for model training. As discussions evolve, so too does the debate surrounding "fair use" and its application, particularly when copyrighted material is sourced from dubious origins.

Adding to the complexities of this situation, the broader creative industry faces the potential ramifications of AI systems trained on improperly sourced content. In March 2025, The Atlantic highlighted the ethical dilemmas Meta encountered during the development of the Llama 3 model. By prioritising expansive datasets such as LibGen, the company has ignited a heated discussion on the appropriateness of exploiting such resources while simultaneously undermining the very authors whose works fuel the AI’s capabilities. Authors argue that the distinction between leveraging public domain resources and appropriating copyrighted works must be more clearly defined, as the integrity of creative expression hangs in the balance.

The legal troubles have continued to mount—April 2025 saw a significant challenge as a U.S. District Judge permitted Zuckerberg to be deposed in ongoing lawsuits. This ruling has provided authors a glimmer of hope that accountability might be forthcoming amid towering corporate interests. It marks a pivotal moment in what could be the first major test of the boundaries of fair use in AI training, especially involving material obtained through piracy. Several lawsuits have emerged not just from U.S. authors but also from French publishers and trade groups, who are asserting similar claims about unauthorized use of their works.

However, while the outcry against Meta's actions is palpable, there exists a complex, sometimes contradictory perspective emerging from certain sectors of the writing community. Some authors have found a perverse relief in the acknowledgment that their works have garnered enough attention to be included in such high-profile AI training datasets. The sentiments echo a broader tension where creators grapple with their content being misappropriated against the backdrop of the undeniable demand for accurate, well-curated training data for AI systems.

As authors reflect on their own experiences, many express ambivalence about AI's role in disseminating information. The crux of their argument rests on the distinctions between reference works and more creative literary pieces. While one author contemplates their reference books being manipulated into AI models, preferring this outcome over the potential for the AI to generate inaccurate information, they simultaneously warn against the risks posed by exploiting creative works without due recognition or compensation. They draw a firm line, asserting that while informative texts might withstand such scrutiny, original novels and artistic expressions deserve greater respect protection against what they describe as "identity theft."

The unfolding landscape suggests an impending cultural reckoning; one where the industries surrounding publishing and digital innovation must inextricably link responsibility with technological advancement. As tech leaders continue to tout the latest AI capabilities, the question remains: how can we evolve in a manner that respects the myriad voices and works that collectively define our cultural heritage? The legal battles ahead might not only shape AI's role in creative industries but could also redefine the foundational principles surrounding content creation and intellectual property for a digital age that continues to push boundaries.

## Reference Map:

* Paragraph 1 – [[1]](https://www.creativebloq.com/ai/call-me-a-traitor-but-i-dont-mind-ai-being-trained-on-my-content-even-after-meta-stole-my-book), [[2]](https://www.reuters.com/technology/artificial-intelligence/meta-knew-it-used-pirated-books-train-ai-authors-say-2025-01-09/)
* Paragraph 2 – [[2]](https://www.reuters.com/technology/artificial-intelligence/meta-knew-it-used-pirated-books-train-ai-authors-say-2025-01-09/), [[4]](https://www.ft.com/content/b1f4965f-6ea6-4afd-968b-76bb2f7acfc4)
* Paragraph 3 – [[3]](https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/?utm_source=apple_news), [[4]](https://www.ft.com/content/b1f4965f-6ea6-4afd-968b-76bb2f7acfc4)
* Paragraph 4 – [[5]](https://apnews.com/article/df4dec4aef8924d38d258212e0654a3d), [[6]](https://www.reuters.com/technology/artificial-intelligence/french-publishers-authors-file-lawsuit-against-meta-ai-case-2025-03-12/)
* Paragraph 5 – [[1]](https://www.creativebloq.com/ai/call-me-a-traitor-but-i-dont-mind-ai-being-trained-on-my-content-even-after-meta-stole-my-book), [[3]](https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/?utm_source=apple_news)

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.creativebloq.com/ai/call-me-a-traitor-but-i-dont-mind-ai-being-trained-on-my-content-even-after-meta-stole-my-book> - Please view link - unable to able to access data
2. <https://www.reuters.com/technology/artificial-intelligence/meta-knew-it-used-pirated-books-train-ai-authors-say-2025-01-09/> - In January 2025, authors including Ta-Nehisi Coates and Sarah Silverman filed a lawsuit against Meta Platforms, alleging the company used pirated books from the LibGen database to train its AI systems. Internal Meta documents revealed awareness of the pirated status of the works, leading to claims that Meta knowingly utilized the dataset for its Llama AI model. The case highlights the ongoing debate over AI companies' use of copyrighted material without proper authorization.
3. <https://www.theatlantic.com/technology/archive/2025/03/libgen-meta-openai/682093/?utm_source=apple_news> - In March 2025, The Atlantic reported that Meta faced an ethical dilemma while developing its AI model, Llama 3. The company opted to use pirated resources, notably the Library Genesis (LibGen) database, which offers over 7.5 million books and 81 million research papers. This decision, backed by Meta CEO Mark Zuckerberg, is now part of a copyright infringement lawsuit involving several authors. The article discusses the implications of using pirated content for AI training and the broader impact on the creative industry.
4. <https://www.ft.com/content/b1f4965f-6ea6-4afd-968b-76bb2f7acfc4> - In April 2025, the Financial Times reported on a significant legal challenge Meta faced, defending itself against a lawsuit filed by a group of U.S. authors, including Ta-Nehisi Coates and Richard Kadrey. The authors allege that Meta unlawfully used copyrighted books from the LibGen database to train its Llama AI models. The case is one of the first major tests to determine whether AI companies can claim 'fair use' of copyrighted material for model training, especially when sourced from pirated databases.
5. <https://apnews.com/article/df4dec4aef8924d38d258212e0654a3d> - In July 2024, the Associated Press reported that Meta CEO Mark Zuckerberg would be deposed in a lawsuit brought by authors, including Sarah Silverman, accusing Meta of using their copyrighted works without consent to train its AI. U.S. District Judge Thomas Hixson denied Meta's attempt to prevent Zuckerberg's deposition, citing evidence of his close involvement in AI initiatives. The class action claims Meta illegally downloaded digital copies of the authors' books, marking a significant development in AI-related copyright disputes.
6. <https://www.reuters.com/technology/artificial-intelligence/french-publishers-authors-file-lawsuit-against-meta-ai-case-2025-03-12/> - In March 2025, French publishers and authors filed a lawsuit against Meta, accusing the company of using their copyrighted works without permission to train its artificial intelligence (AI) model. Three trade groups, including the National Publishing Union, the National Union of Authors and Composers, and the Societe des Gens de Lettres, initiated the legal action in a Paris court. They allege that Meta engaged in unauthorized use of their works to develop generative AI capabilities for its platforms, including Facebook, Instagram, and WhatsApp.
7. <https://www.ibpa-online.org/news/697191/IBPA-Statement-on-Metas-Use-of-Pirated-Books-to-Train-AI.htm> - In March 2025, the Independent Book Publishers Association (IBPA) released a statement supporting the Authors Guild's response to Meta's use of pirated books to train its AI models. The IBPA condemned Meta's decision to use pirated content instead of engaging in ethical licensing agreements with authors and publishers. The association emphasized the importance of compensating rightsholders and criticized Meta's reliance on the 'fair use' defense, arguing that companies like Meta can afford to pay for the content they use.