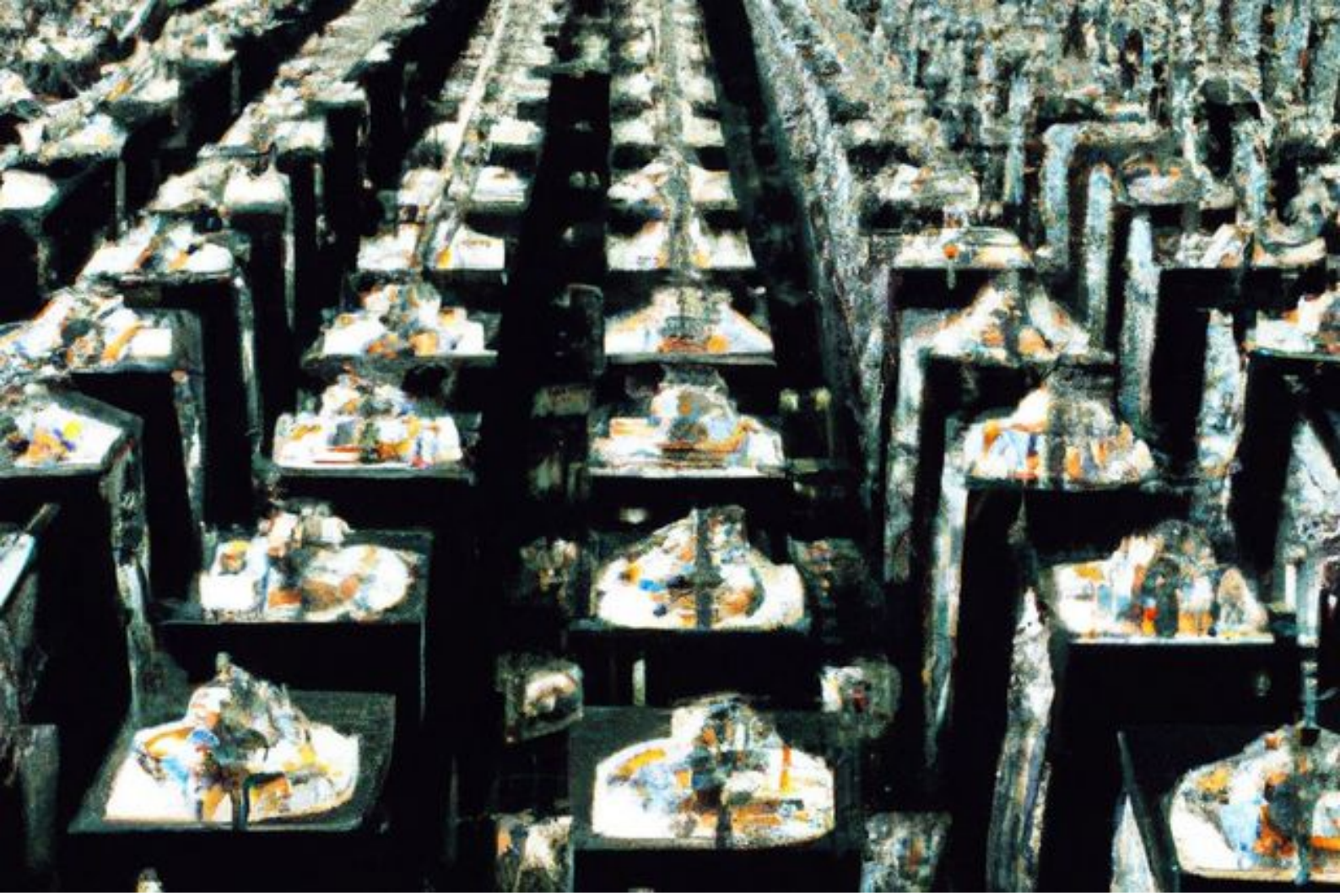


News Featured Stories

Book Authors Take Action Against Nvidia Over AI Data

After Open AI, writers are now going against Nvidia over copyright violations in training the NeMo system LLM

A group of prominent novelists have filed a lawsuit against chip maker Nvidia, alleging the company violated their copyrights by using pirated ebooks to train large language models without permission.

The lawsuit, filed in a US district court in San Francisco, claims Nvidia's AI platform NeMo was trained using a dataset called Books3 that contained nearly 200,000 pirated books copied from an online repository called Bibliotek, a shadow library of pirated books online.

Authors Abdi Nazemian, Brian Keene, and Stewart O'Nan are named as plaintiffs in the proposed action. They argue that by training models on their copyrighted works without a license, Nvidia infringed on their exclusive rights under US copyright law.

The Books3 dataset was initially shared online on hugging face before being removed last year after copyright concerns. However, the lawsuit alleges Nvidia had already obtained copies and continued using the data to train AI models like NeMo that can generate human-like text outputs.

Authors' lawyers successfully fought off a motion to dismiss similar copyright claims against OpenAI over its use of books to train ChatGPT. They aim to secure damages from Nvidia and prevent further distribution of models trained on the pirated material.

Transparency around training data has become a key issue, with some firms limiting disclosures after lawsuits. Experts urge clear standards to balance innovation with rights protection as the AI revolution unfolds.

--
Image credit: A Generative AI self-portrait by DALL·E. Via Wikimedia Commons.

Co-funded by
the European Union