The latest AI copyright lawsuit involves Mike Huckabee and his books

Former Arkansas Governor Mike Huckabee is among a group of authors suing Meta, Microsoft, and other companies over the use of their work in building AI tools.

In a lawsuit filed Tuesday, Huckabee and other authors including Christian writer Lysa TerKeurst allege that their books were pirated and used in datasets that trained AI models. EleutherAI, an artificial intelligence research group, is also named in the suit, as is Bloomberg.

The proposed class action suit is the latest example of authors alleging tech companies used their work without permission to train generative AI models. Over the past several months, a string of popular authors including George R.R. Martin, Jodi Picoult, and Michael Chabon have sued OpenAI for copyright infringement.

The Huckabee case centers on a controversial trove of data called “Books3” that contains more than 180,000 works that are part of the dataset used to train large language models. In August, *The Atlantic* published a searchable database of all the titles in Books3 with author information. Books3 is part of a larger mountain of data called the Pile, created by EleutherAI, that the suit says was used by companies to train their products.

“[Meta and Microsoft] were able to incorporate sophisticated datasets, which included the pirated copyright-protected materials in Books3, as part of the LLM’s training process,

without having to compensate the authors,” the suit reads.

Microsoft declined to comment for this story. Meta, Bloomberg, and EleutherAI didn’t respond to requests for comment.

AI companies rely on massive amounts of public data to train AI models — not just books but also photographs, art, music, and more. As tools like ChatGPT or Stable Diffusion have become easily accessible, there’s been heated debate (and lots of legal action) about how people who provide that data should be compensated. In January, Getty Images sued the company behind AI art tool Stable Diffusion, claiming it unlawfully copied millions of copyrighted images to train its model.

https://www.theverge.com/2023/10/19/23924246/mike-huckabee-books-lawsuit-generative-ai-microsoft-meta-bloomberg-eleutherai-books3