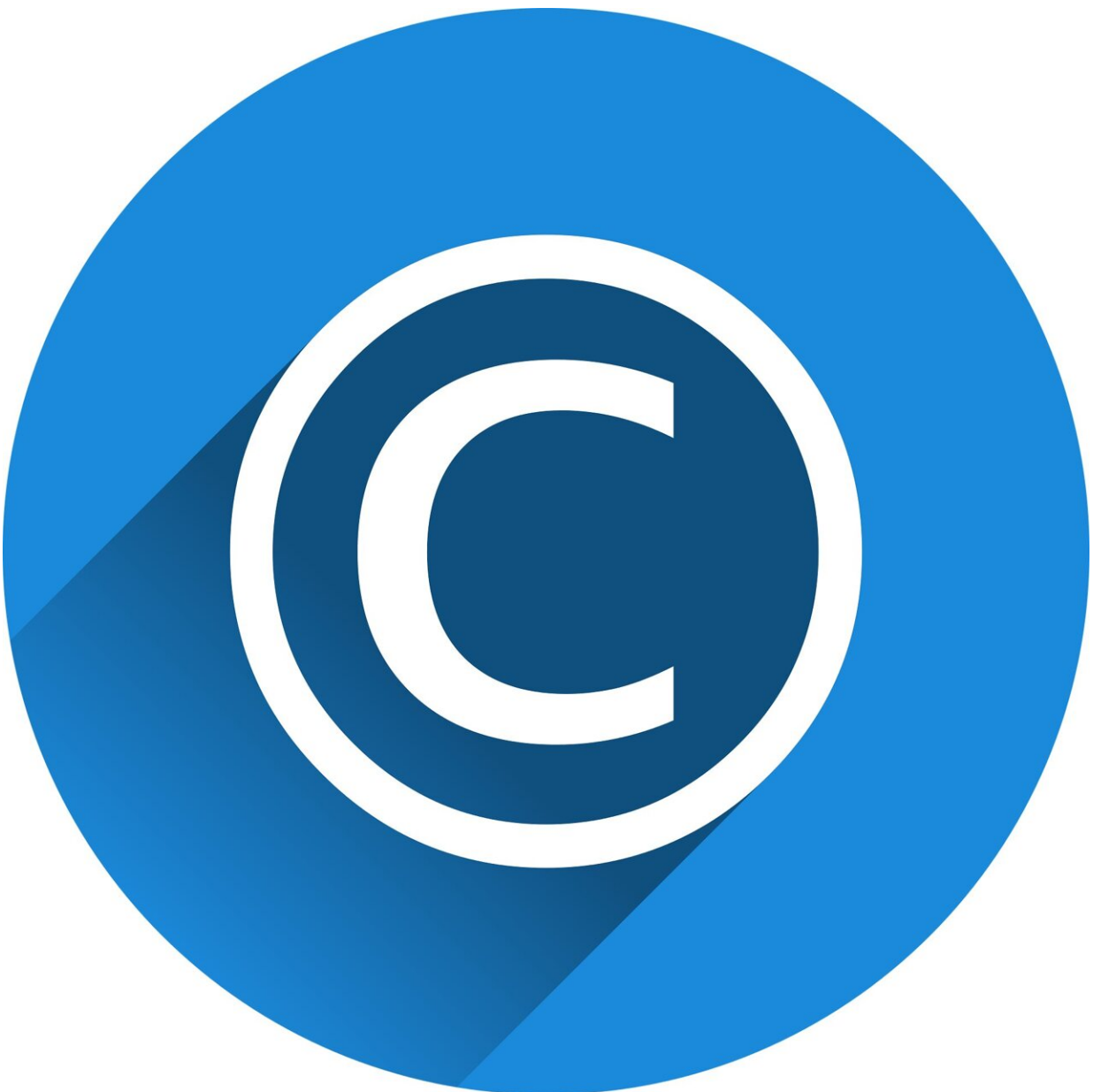


# **Thousands of pirated Australian books—in the age of AI, is copyright law still fit for purpose?**

September 29 2023, by Dilan Thampapillai

---



Credit: Pixabay/CC0 Public Domain

Thousands of Australian books [have been found](#) on a pirated dataset of ebooks, known as Books3, used to train generative AI. Richard Flanagan, Helen Garner, Tim Winton and Tim Flannery are among the leading local authors affected—along, of course, with writers from around the world.

A [search tool](#) published by [The Atlantic](#) makes it possible for authors to find out whether their books are among the nearly 200,000 in the Books3 dataset.

Many of these writers have reacted angrily about their works being included in these datasets without their knowledge or consent. Flanagan [told the Guardian](#), "I felt as if my soul had been strip mined and I was powerless to stop it".

"Turning a blind eye to the legitimate rights of [copyright](#) owners threatens to diminish already-precarious creative careers," said Olivia Lanchester, chief executive of the Australian Society of Authors, in [an official response](#) this week.

## **AI moving at speed**

Authors have turned to [copyright law](#) because it is the body of law that has traditionally protected authors and other creators from the appropriation of their works.

However, laws designed for the pre-AI era have little meaning in the

post-OpenAI world.

Just last year, the issue of AI was only faintly on the cultural radar. But while AI technology is moving at high speed, the law moves slowly.

It took a very significant amount of time for copyright law to first appear. The first copyright law, the [Statute of Anne](#), emerged in 1710 after protracted lobbying by stationers (publishers).

In a more modern context, it took 20 years from the time Australian courts first recognized a system of Aboriginal law existed, with the [Milirrpum decision](#) in 1971—meaning terra nullius was implausible—to the High Court handing down the landmark Mabo decision that erased terra nullius, in June 1992. In the interim, injustice reigned.

The question that now confronts us is whether we can wait for the law to catch up with the rapid advances of technology—or whether we must jumpstart the process.

## **A spate of copyright disputes**

There has been a spate of copyright disputes around AI datasets and copyright-protected works.

Earlier this month, the US Authors Guild [filed a class action](#), with 17 authors including Jonathan Franzen and Jodi Picoult, against OpenAI for copyright infringement.

This followed the first copyright lawsuit against OpenAI in July. It was filed by authors Mona Awad and Paul Tremblay, for using their books to train its AI, ChatGPT, without their consent.

And in August, Benji Smith was forced to take down his website

Prosecraft, which used an algorithm to trawl through more than 25,000 books (again, without authors' consent) to produce analysis designed to give writing advice.

## **Copyright is not the answer**

While it's true that the uploading of works into a dataset is an act of copyright infringement, that only pertains to a one-off act of infringement.

No doubt, the liability would be large if thousands of works were involved and thousands of authors were to sue (as with the US Authors Guild [class action](#)), but the damages obtained by an individual author would be relatively small, making it not worth suing. The large commercial interests driving the development of the datasets and related AI tools are likely to withstand these lawsuits even if they are found liable.

Likewise, copyright law's rules on fair dealing in Australia and fair use in the United States would likely protect some uses.

Further, the outputs from AI that have been trained on these datasets are not likely to result in works that satisfy the substantial similarity threshold (which means that when the two works are compared side by side, they must be similar) for copyright infringement in most jurisdictions, including Australia.

## **'A type of market failure'**

Copyright law has previously had to balance the interests of creators with those of technology developers.

This happened when the photocopier was invented, when video cassette recorders were developed, when blank tapes became widely available and when peer-to-peer [copyright infringement](#) took off during the digital era.

The difference then was that these technologies did not fundamentally threaten artistic and creative labor in the way AI does.

To appropriate a part of someone's market is a radically different thing to producing a product that could entirely displace them in that market.

Yet this is the direction we're heading in. And it requires a very significant rethink about the regulation of technology.

A type of market failure is occurring here, because authors are not being compensated even though their works, collectively, are the basis for new and commercially viable AI products.

When the sale of blank tapes began, [the government responded](#) with a levy on every blank tape sale, which sent money back to copyright owners.

Something like the blank tape levy might need to be considered for AI. This would mean every time somebody uses an OpenAI-type tool for which they pay a fee, some small portion of the fee would revert to copyright owners.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Thousands of pirated Australian books—in the age of AI, is copyright law still fit for purpose? (2023, September 29) retrieved 9 May 2024 from <https://techxplore.com/news/2023-09-thousands-pirated-australian-booksin-age.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.