putting
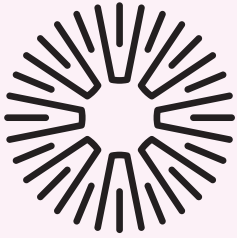# the ai_

in Fair Use

Authors' Abilities to Promote Progress

AUTHORS ALLIANCE

Authors Alliance advances the interests of authors who want to serve the public good by sharing their creations broadly. We create resources to help authors understand and enjoy their rights and promote policies that make knowledge and culture available and discoverable.

All underlined phrases have citations, which can be found in our digital bibliography. Access it with the QR code.

Authors v. GenAI?

For authors like Michael Chabon, Ta-Nehisi Coates, and Jacqueline Woodson, generative artificial intelligence (GenAI)—tools like Open AI's ChatGPT and Google's Gemini—is not just innovative novel technology; it uses their books without credit, consent, or compensation. These authors, like others in their field, are concerned that such uses allow GenAI developers to benefit from authors' works while many authors struggle to make a living.

In their class action lawsuit challenging OpenAI, these and other authors argue that OpenAI infringes their copyrights by using their books to "train" ChatGPT. OpenAI's response to similar lawsuits is that using copyrighted works as "training data" is fair use, a legal doctrine that allows copyrighted works to be used without permission in select statutory situations. Courts are still deciding who is right in the context of GenAI, so the question stands: **Is using copyrighted works as training data fair use?**

Fair Use

Fair use is a multi-factor standard that allows limited use of copyrighted material—even without credit, consent, or compensation.

Four Fair Use Factors

1

The Purpose and Character of the Use

2

The Nature of the Copyrighted Work

3

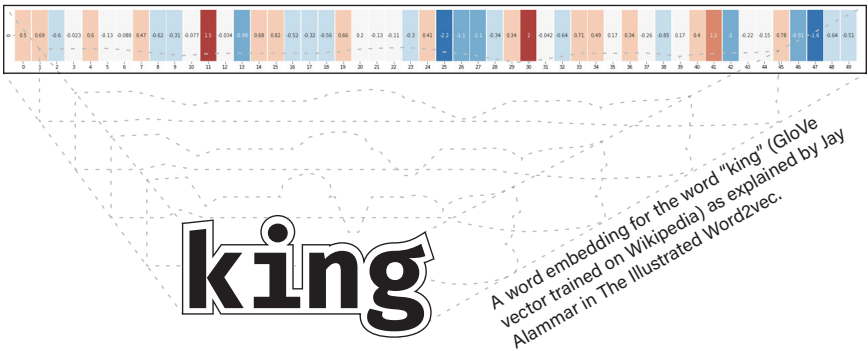**The Amount or Substantiality of the Portion Used**

4

The Effect of the Use on the Potential Market for or Value of the Work

The Supreme Court has said that "[t]he four statutory factors are to be explored and weighed together in light of copyright's purpose." No factor is determinative of the test's outcome, but particular emphasis is placed on factor 1 and factor 4. While no US court has expressly endorsed the unauthorized use of copyrighted works as training data for GenAI, those developers might be right legally.

GenAI Explained

GenAI relies on training data, but some authors still view that use as leveraging consentless copies of their work. Though it seems like a GenAI model may "read" the training data to teach itself about human language patterns, scholars have shown that the process is more **abstract** and **mechanical** than that. GenAI models transform each word into many different numbers and use those numbers to predict words that together make an algorithmic output.

A word embedding for the word "king" (GloVe vector trained on Wikipedia) as explained by Jay Alammar in The Illustrated Word2vec.

We believe that this transformative process (among other factors) makes training data fair use. But we recognize that GenAI developers' unauthorized use of copyrighted works does not seem "fair" to some people. However, there is an important benefit to these uses; using high-quality training data promotes "**progress**," the U.S. Constitution's justification for copyright law.

Progress and GenAI

Historically, courts and legal scholars have conceptualized "progress" as just prioritizing financial incentives to promote the efficient production of more works. But progress can be conceptualized more broadly, both legally and sociotechnically.

[The Congress shall have Power…] To promote the Progress of Science and useful Arts, by securing for limited Times to Authors…the exclusive Right to their respective Writings and Discoveries.

Patent and Copyright Clause
U.S. Constitution, Article I, Section 8, Clause 13

Progress is also about, as former U.S. Copyright Office Register and a leading architect of the Copyright Act of 1976 Barbara Ringer put it, "the widest possible access to information of all kinds" while promoting authors' incentives and preserving their economic rights over their works. The fair use doctrine strikes this balance.

Legally, "progress" as applied to GenAI means that training data can be fair use when it helps create new information that creators can make freely and publicly accessible. Authors should consider proactively sharing their works as training data to promote four values that stem from widely accessible works: **equity**, **education**, **enjoyment**, and **efficiency**.

Advancing Equity

Sharing helps GenAI produce less biased outputs, better serving the public and marginalized communities.

Amplifying Education

Sharing helps enable well-informed content that enhances public knowledge.

Allowing Enjoyment

Sharing helps authors leverage GenAI to create more fun and innovative works.

Augmenting Efficiency

Sharing helps aid authors and creatives in producing new works more quickly by accelerating brainstorming and research.

Danger of Biased Training Data

"AI needs good data. If the data is incomplete or biased, AI can exacerbate problems of bias."

**National Science and Technology Council**
Obama Administration

The danger of biased training data is that it could generate biased outputs. Take Wikipedia, which is a massive Creative Commons-licensed project. Former Executive Director of the Wikimedia Foundation Katherine Maher stated that almost every AI system uses facts from Wikipedia as training data. But only 8.5% of Wikipedia editors are women, which manifests as gender bias in Wikipedia articles, from length to tone to existence. Octavia Butler, who published her acclaimed iconic novel *The Parable of the Sower* in 1993, did not have a Wikipedia article until 2003. The article's substance was a single sentence. Imagine how a GenAI system trained on these articles could reflect some gender bias.

Using Wikipedia as training data exemplifies what Professor Amanda Levendowski dubbed "biased, low-friction data"(BLFD) because Creative Commons (CC) licenses often grant the public permission to use the work, reducing their friction for use as training data. But, as seen with the potential gender bias in Wikipedia articles, CC-licensed and other works can amplify biases.

Case Study: BLFD

While BLFD is easy to use, biased data can create major problems, especially for people with marginalized identities.

Google's word2vec is a model integrated into machine learning (ML) algorithms that recognizes relationships between words. While word2vec may be a helpful tool for Google to use to improve its GenAI and AI technology, it also reveals the **gender and racial biases** of the data used to train it. Researchers found that word2vec associated **men** with **computer programmers** and **women** with **homemakers**. The tool also associated **Black men** with the word **assaulted** while associating **white men** with the phrase "**entitled to**." Many legal and sociotechnical scholars have found that gender and racial biases in similar models is an **evergreen** problem.
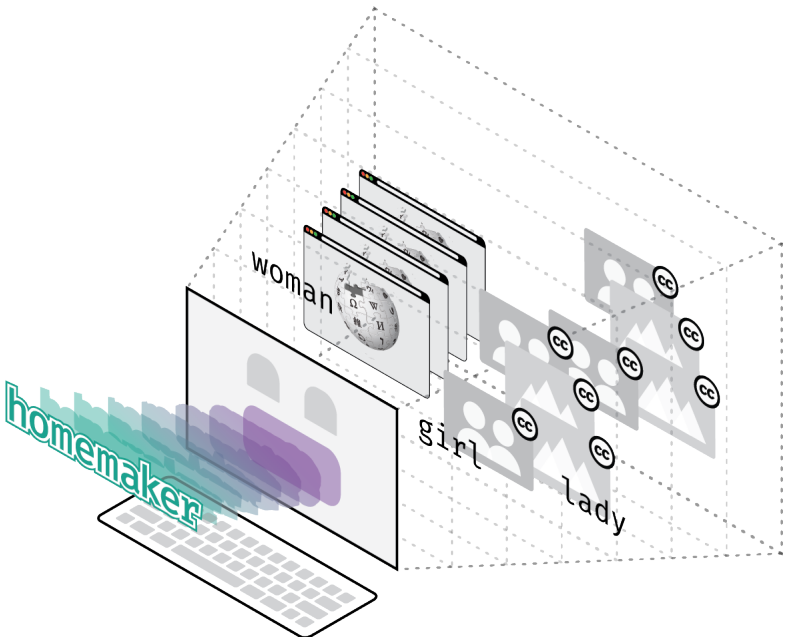
Intersectional Lens, Intersectional Harm

These Wikipedia and Word2vec examples illustrate GenAI's capacity to produce outputs that exact intersectional harm. Coined by Professor Kimberlé Crenshaw, intersectionality blends critical theories, including Critical Race Theory and feminism, to create:

> "[A] lens...for seeing the way in which various forms of inequality often operate together and exacerbate each other. We tend to talk about race inequality as separate from inequality based on gender, class, sexuality, or immigrant status. What's often missing is how some people are subject to all of these, and the experience is not just the sum of its parts."

Legal and sociotechnical scholars–including Dr. Joy Buolamwini and Dr. Timnit Gebru, Professor Sonia Katyal and Jessica Y. Jung, and Dr. Safiya Umoja Noble–have adopted an intersectional lens to interrogate bias in AI systems like face surveillance, body scanners, and search engines.

The Nuance of the Implicit Bias Problem

There is no convenient or complete solution to solve GenAI's implicit bias problem. Some scholars even believe that the inclusion of marginalized people to "mitigate" bias can normalize their oppression.

"More perniciously, inclusion reinforces the structural sources of violence it supposedly addresses."

Dr. Anna Lauren Hoffmann
University of Washington Information School

By quantifying human expression and authors, other AI systems like face surveillance can separate marginalized people from what makes them human—their character, creativity, and curiosity.

On the other hand, GenAI is a creative tool that can also amplify the experiences of marginalized people through diverse training data. Data shifts cannot fully disrupt or dismantle the harms faced by marginalized people, but some scholars have suggested that broad access to copyrighted works may be a necessary step toward reducing bias.

Dominant Players Concentrate Power

GenAI platforms and datasets are concentrated and controlled by the most powerful institutions. The Microsofts and Metas of the world can acquire copyrighted works as training data in 4 ways.

Build It

Sweep up users' copyrighted works through terms of service (Meta used public Facebook and Instagram posts to train MetaAI; Google used Gmail to train an AI model).

Borrow It

Scrape copyrighted works from the Internet without permission (Chabon and other authors challenge this practice).

Buy It

License copyrighted works (OpenAI has offered news publishers between $1 million and $5 million for use).

Bypass It

Use works with limited or no copyright restrictions (such as CC-licensed or public domain works).

Powerful players can build it and buy it. With structural support from Microsoft, OpenAI can also pursue the legally untested borrow it model. This is a competitive advantage for established companies over upstarts, which may experience friction in training AI with copyrighted works. A less-biased upstart might never compete financially or logistically with the incumbents.

Why Diverse Training Data?

Diverse training data combats AI biases by...

Minimizing Misinformation

GenAI datasets trained on lower-quality works can inadvertently undercut **education** through misinformation, like hallucinations. Famously, Google's chatbot Bard incorrectly stated that the James Webb Space Telescope took the first images of a planet outside our solar system when it did not.

Countering Less-Diverse Creators

GenAI datasets trained on BLFD can amplify the views of a less diverse group of creators. For the **equitable** development of GenAI, authors' works can diversify datasets to improve outputs

But it won't combat them all. E.g., diverse training data **does not**...

Diversify AI Developers

There is an AI developer diversity crisis across race and gender, with 80% of AI professors being men and a fraction of the workforce at top AI firms (e.g., Google and Meta) being people of color. This harms marginalized communities as "[t]here is an intersection between discriminatory workforces and discriminatory technology."

Copyright Cassandras

We believe authors can get comfortable with GenAI platforms using their works to debias AI systems, even as many of these systems currently use copyrighted works as training data without consent. Drawing on legal precedent and empirical work, many legal scholars believe training data is likely to be fair use in some circumstances–like combating bias. Soon, courts may concur.

"[C]opyright law also has the profound power to unbias [AI systems]…Fair use can, quite literally, promote creation of fairer AI systems."

Professor Amanda Levendowski
Georgetown Law

"Restricting the training data for LLMs to public domain and open license material would tend to encode the perspectives, interests, and biases of a distinctly unrepresentative set of authors."

Professor Matthew Sag
Emory Law

Dr. Mehtab Khan
Harvard Law

Professor Ben Sobel
Cornell Tech

"Accessing copyright work may also be necessary for the purpose of auditing, testing, and mitigating bias in datasets… Here, it may be useful to rely on the *flexibility of fair use*, and support access for researchers and auditors" of AI.

"If engineers made unauthorized use of copyrighted data for the sole purpose of debiasing an expressive program,…fair use would excuse it."

Dr. Alex Hanna
Distributed AI Research Institute

What Authors Can Do

Authors can further the progress of GenAI by sharing their works as unencumbered training data. By bolstering the bypass it model, authors' works may help curb AI bias.

Actions Authors Can Do To Promote Progress

1

Contract Negotiations

Signing away all of an author's rights may result in less use of their work, so authors can advocate for favorable terms that ensure wider access to their works.

2

Open Access

Authors can increase the reach of their works by employing open access publishing, which empowers authors to freely disseminate their works while maintaining certain rights. Authors can consider obtaining a Creative Commons license, which creates flexible open access licenses.

3

Rights Reversion

Authors can request their rights back, given that certain conditions are met, to widen their works' reach.

4

Termination of Transfer

Authors can regain rights to their works after at least 35 years, allowing them to make their works more widely accessible on their own terms.

Was this zine too much? ChatGPT sums it up:

"In the ever-evolving saga of AI versus authors, it's a comedy of errors where fair use meets fair game. With legal battles, moral dilemmas, and the occasional plot twist, authors are the unlikely heroes of a digital sitcom. But fear not, dear readers, for as they navigate the labyrinth of copyright chaos, armed with legal savvy, they remind us that in this high-tech comedy, progress is always the punchline. So here's to the authors – may they choose their rights wisely, may their pens stay sharp, and may their stories continue to shape the future of AI, one laugh at a time!"

Thanks ChatGPT (with client consent) for the TLDR.

This zine was created by Intellectual Property and Information Policy (iPIP) Clinic student attorneys Mariah Johnson and Marcus Liou practicing pursuant to DC App. R. 48 under the supervision of Amanda Levendowski