

Does Generative AI Infringe Copyright?

James Grimmelmann
Cornell { Tech | Law School }

Trustworthy AI
October 27, 2025

Katherine Lee, A. Feder Cooper, & James Grimmelmann,
*Talkin' 'Bout AI Generation: Copyright and the Generative AI
Supply Chain*, 72 Journal of the Copyright Society of the
U.S.A. 251 (2025)



Katherine Lee



A. Feder Cooper

(1) GETTY IMAGES (US), INC (a company incorporated under the laws of the	J. DOE 1 and J. DOE 2, individually and on behalf of all others similarly situated,	§., P.M., N.G., R.F., J.D. and G.R., and on behalf of all others similarly
(2) GETTY IMAGES INTERNATION (a company incorporated under the law	Individual and Representative Plaintiffs,	
(3) GETTY IMAGES (UK) LIMIT (4) GETTY IMAGES DEVCO UK LI	v.	Plaintiffs,
(5) ISTOCKP SARAH SILVERMA (a company incorporated CHRISTOPHER GO	THE NEW YORK TIMES COMPANY	Plaintiff,
(6) THOMAS M. B RICHARD KADREY (a company incorporated under th	v.	GLE DEEPMIND, and
	MICROSOFT CORPORATION, OPENAI, INC., individual; OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, individual; OPENAI OPCO LLC, OPENAI GLOBAL LLC, N, an individual; OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC, presentative Plaintiffs,	
-al v.		
PAUL TREMBLAY, an indi MONA AWAD, an individu	OPENAI, INC., a D Delaware limited par limited liability corp Indi limited liability company; OPENAI STARTU	Defendants.
	PLAINTIFFS P.M., K.S., B.B., S.J., N.G., C.B., S.N., J.P., S.A., L.M., D.C., C.L., C.G, R.F., N.J., and R.R., individually, and on behalf of all others similarly situated,	META PLATFORMS INC a Delaware SARAH ANDERSEN, an individual; KELLY MCKERNAN, an individual; KARLA ORTIZ, an individual,
v.		Individual and Representative Plaintiffs,
OPENAI, INC., a D L.P., a Delaware lim Delaware limited liab Delaware limited liab GP I, L.L.C., a De	Plaintiffs,	v.
STARTUP FUND I, OPENAI LP, OPENAI INCORPORATED, OPENAI STARTUP OPENAI GP, LLC, OPENAI STARTUP FUND I, LP, OPENAI STARTUP FUND GP I, LLC, OPENAI STARTUP FUND MANAGEMENT LLC, MICROSOFT CORPORATION and DOES 1 through 20, inclusive,		STABILITY AI LTD., a UK corporation; STABILITY AI, INC., a Delaware corporation; MIDJOURNEY, INC., a Delaware corporation; DEVIANTART, INC., a Delaware corporation,
		Defendants.

Does generative AI
infringe copyright?

Obviously yes

Copyright in one slide

- “Copyright protection subsists ... in original works of authorship fixed in any tangible medium of expression” 17 U.S.C. § 102(a)
- “[T]he owner of copyright ... has the exclusive right[] ... to reproduce the copyrighted work” 17 U.S.C. § 106(1)

Overview

The Common Crawl corpus contains petabytes of data, regularly collected since 2008.

Choose a crawl... ▾

The corpus contains raw web page data, metadata extracts, and text extracts.

Common Crawl data is stored on Amazon Web Services' Public Data Sets and on multiple academic cloud platforms across the world.

Learn how to [Get Started](#).

Access to the corpus hosted by Amazon is [free](#).

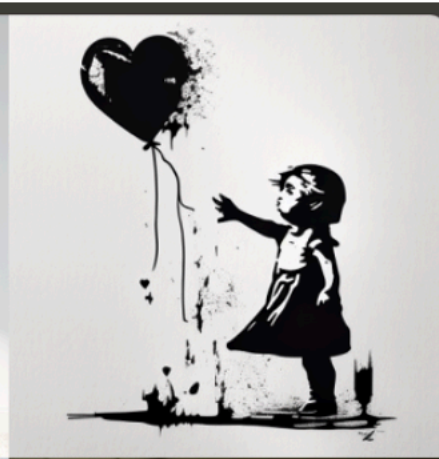
You may use Amazon's cloud platform to run analysis jobs directly against it or you can download it, whole or in part.

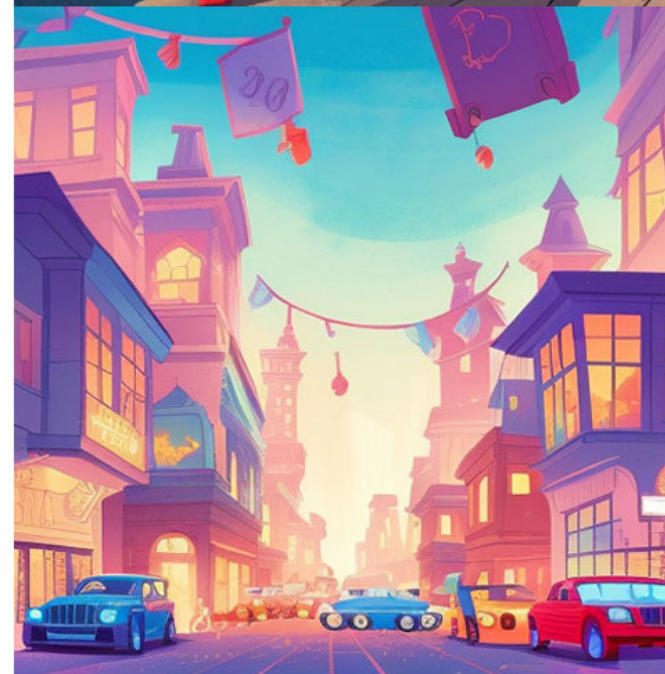
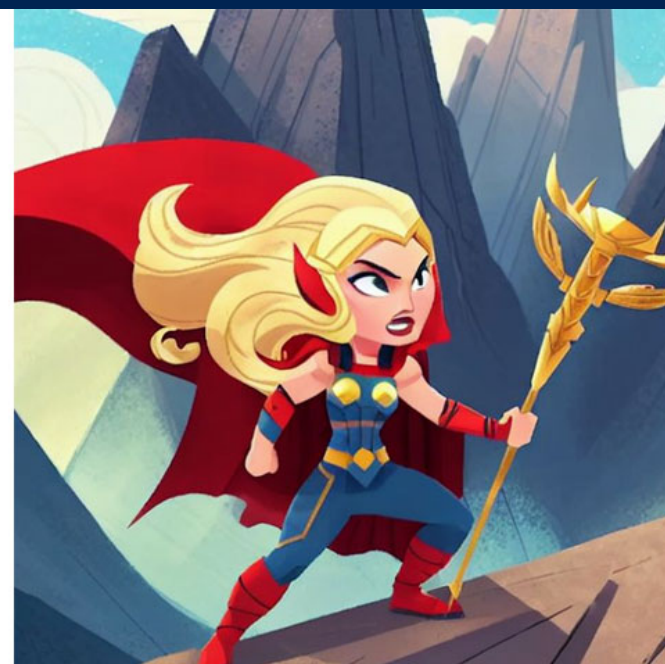
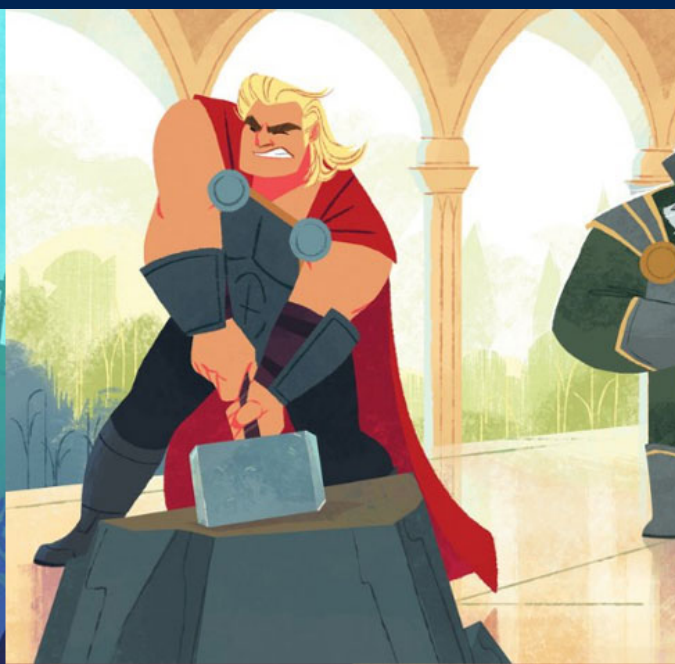
You can search for pages in our corpus using the [Common Crawl URL Index](#).

Check out the [Example Projects](#), view [Use Cases](#), or [Statistics](#) for our crawls.









Copies everywhere

- There are copies of millions of works
 - In the training datasets
 - In the outputs
 - In the trained models (which can be seen “with the aid of a machine or device,” 17 U.S.C. § 101)
- All of these are infringing
- Or I should say, *potentially* infringing ...

Does generative AI
infringe copyright?

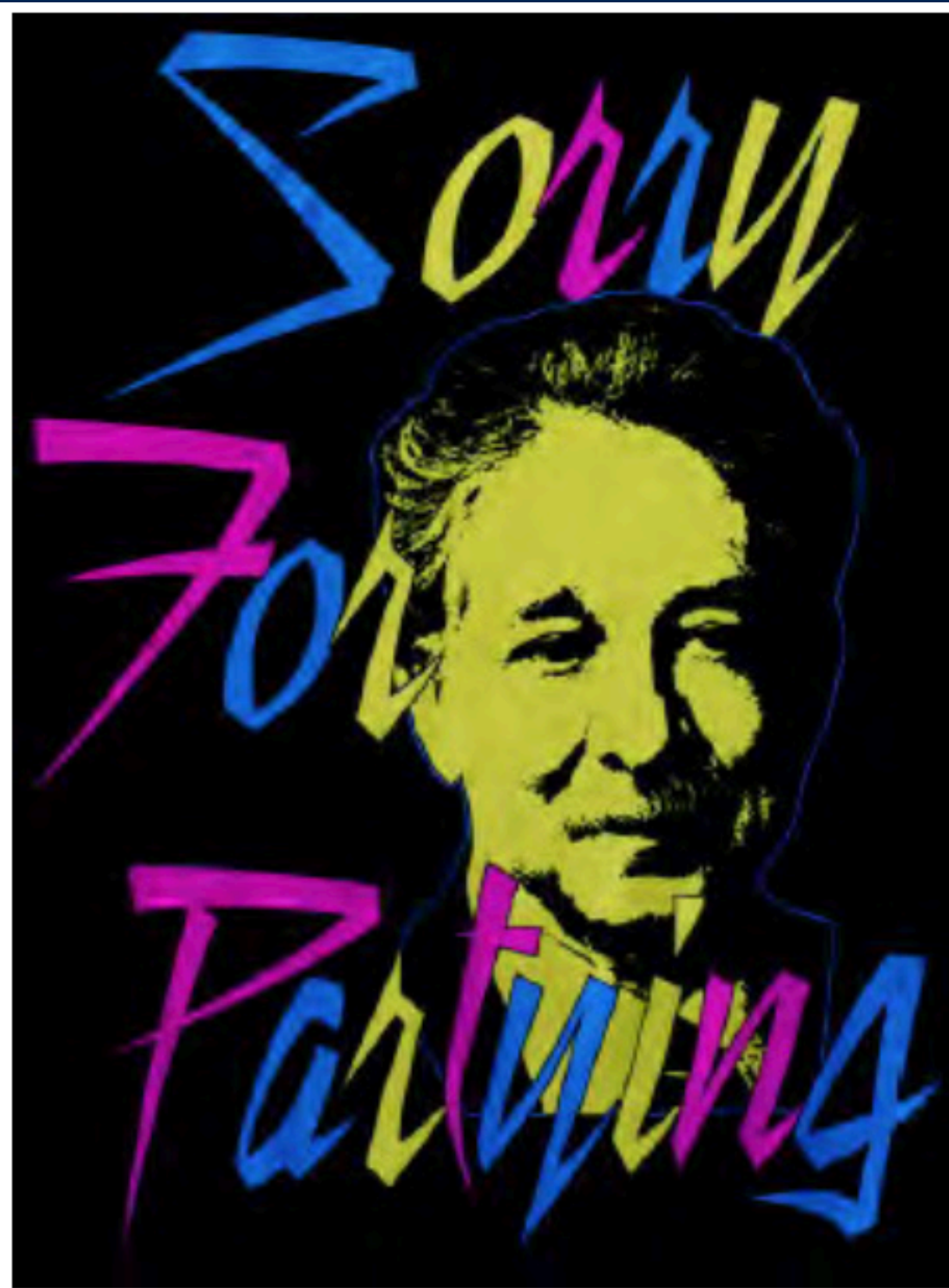
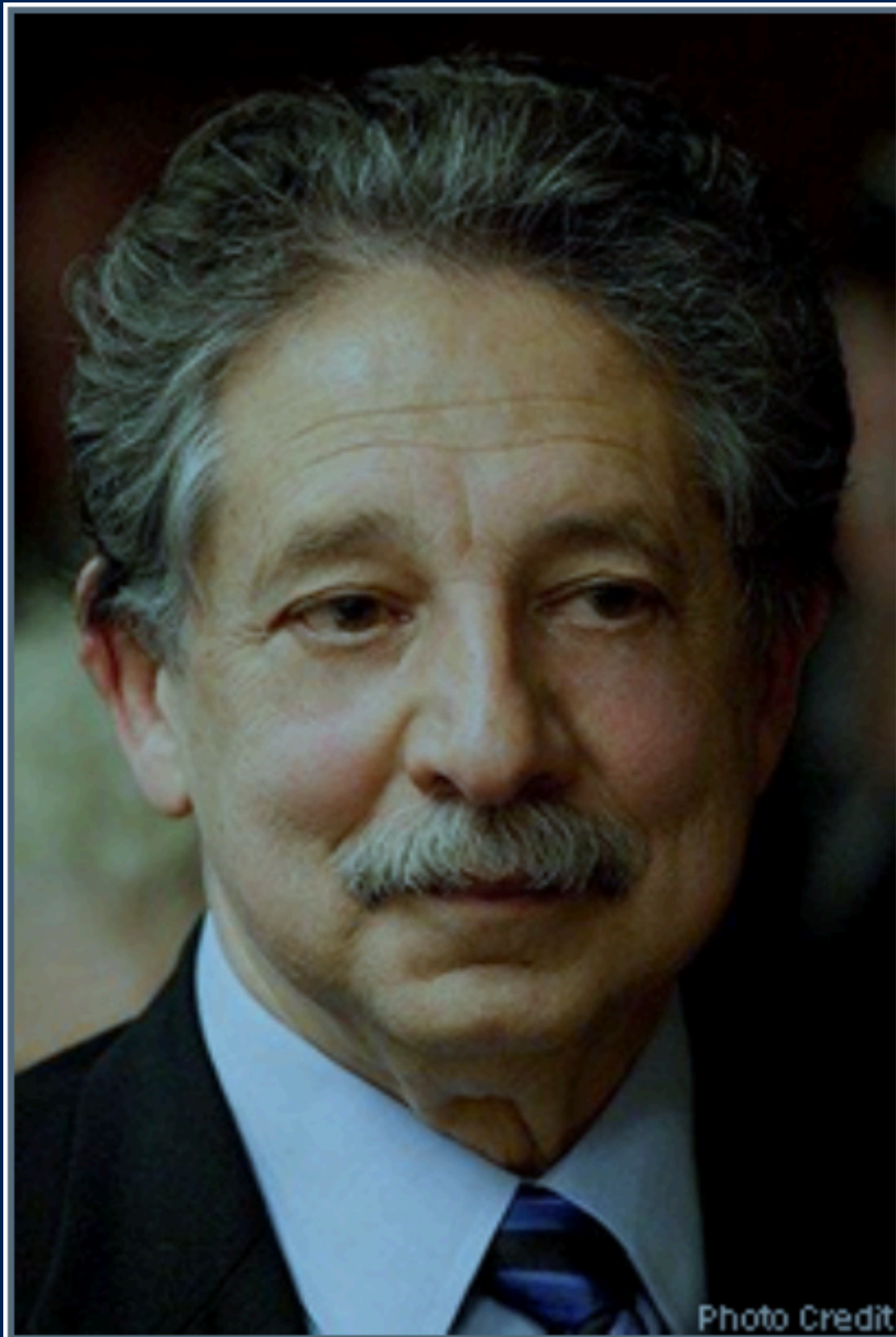
Obviously no

Fair use

17 U.S.C. § 107

“[The fair use of a copyrighted work ... for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. ... [T]he factors to be considered shall include—

- (1) the purpose and character of the use, ...
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used ... ; and
- (4) the effect of the use upon the potential market for or value of the copyrighted work.”



⁹ BHP was founded in Melbourne in 1885. Billiton was founded in 1860. In June 2001, the two companies merged. In the 2003-2004 financial year, it had total revenue of \$34.087 billion, a total market capitalization of \$136.35 billion. It employed ⁴ 35,000 people, making it the second-largest mining group company globally. It was listed on stock exchanges in Australia, London, and New York. The company is the world's third-largest supplier of iron ore. The company has more than 100 years of business relationships with China, including importing and exporting mineral products and steel, minerals, and land and sea oil exploration. Board Chairman Don Argus and CEO Charles Goodyear.

Over the past 100 years, from the trade of resource products, domestic and foreign joint ventures to the procurement of mining machinery and equipment in China, ¹ BHP Billiton has ¹¹ established a long-term and friendly all-round partnership with Chinese business partners. As the world's second-largest economy, China is critical to developing BHP Billiton. Over the ¹⁵ past 40 years, BHP Billiton has supplied China with more than ¹⁴ 2 billion tons of iron ore and coking coal, liquefied natural gas, copper concentrate, nickel, and other resource products in the process of China's rapid industrialization and urbanization, ¹ providing a solid resource guarantee for China's development. Like many multinational companies, BHP Billiton is a vital participant, witness, contributor, and beneficiary of China's reform and opening up and has forged an indissoluble ⁸ relationship with China.

BHP Billiton Group is the world's largest production and distribution enterprise of various

Match Overview

79%

< Currently viewing standard sources >

[EN View English Sources \(Beta\)](#)

Matches

1	Submitted to University... Student Paper	23%	>
2	Submitted to Queensla... Student Paper	11%	>
3	Submitted to RMIT Uni... Student Paper	6%	>
4	Submitted to University... Student Paper	6%	>
5	Submitted to Flinders U... Student Paper	5%	>
6	Submitted to University... Student Paper	4%	>
7	Submitted to Swinburn... Student Paper	4%	>

Google Books Ngram Viewer

Q

artificial intelligence,machine learning,copyright

X

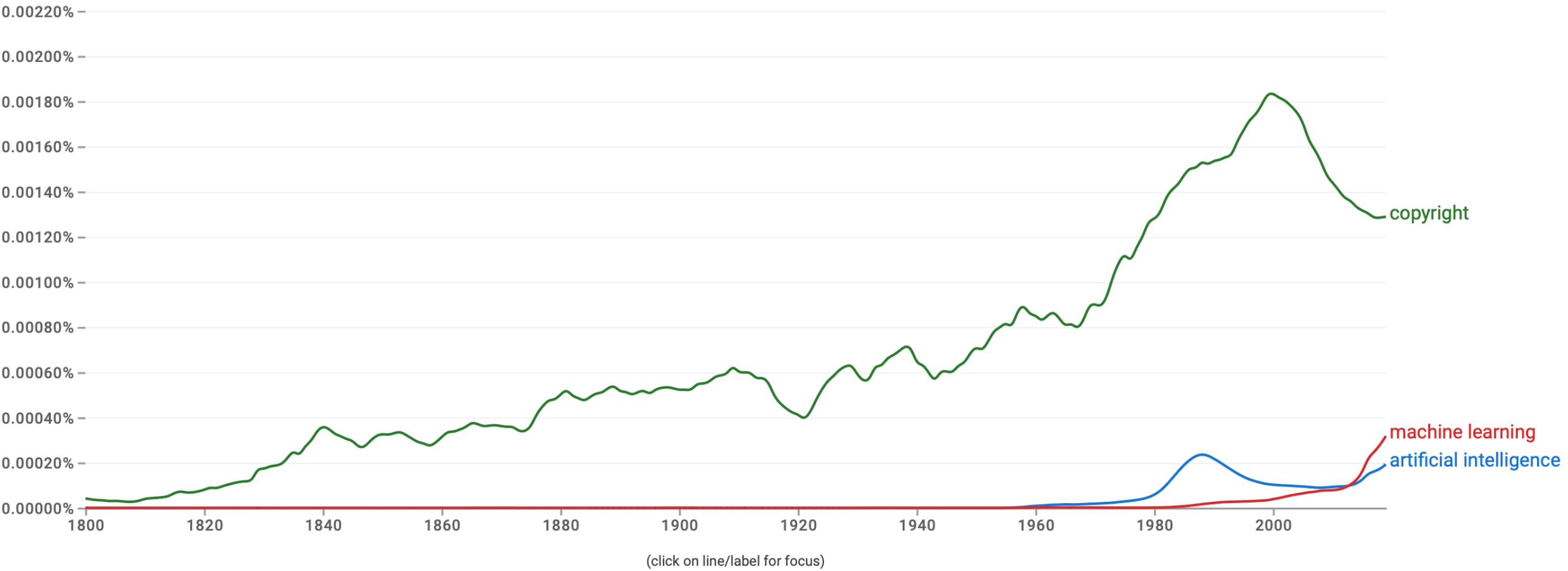
?

1800 - 2019

English (2019)

Case-Insensitive

Smoothing



Fair use everywhere

- Outputs: transformative fair use
- Training data: nonexpressive fair use
- Models: nonexpressive *and* transformative fair use
- ... but what if *some* of the outputs infringe?

Does generative AI
infringe copyright?

Obviously only by users



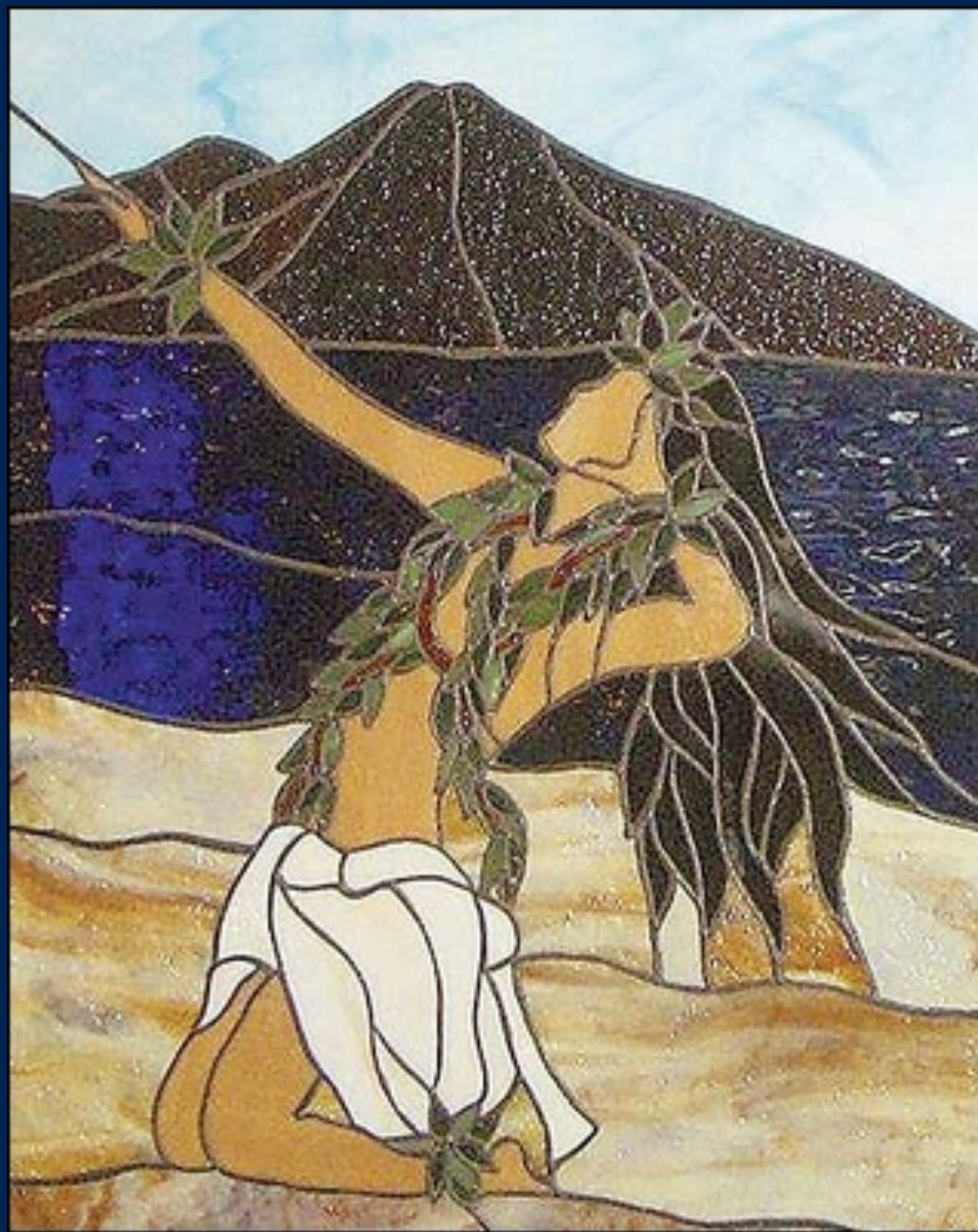


Why not notice and takedown?

- Where does the Copyright Act authorize it?
- How can copyright owners find infringing works?
- How can AI companies block infringing outputs?

Does generative AI
infringe copyright?

It depends



Training Set



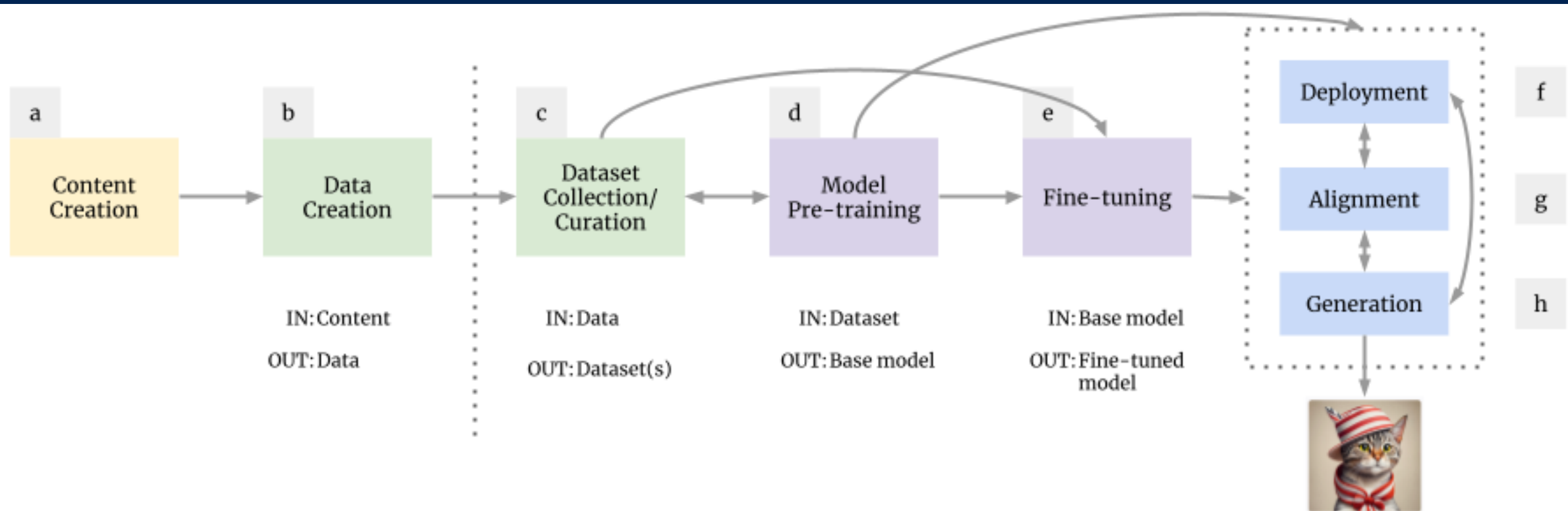
*Caption: Living in the light
with Ann Graham Lotz*

Generated Image



*Prompt:
Ann Graham Lotz*





A diverse supply chain

- A dataset/model/app could ...
 - ... draw on multiple upstream ones
 - ... be used by multiple downstream ones
- Every stage could be ...
 - ... public or private
 - ... proprietary or open
 - ... carried out by the same or different parties

Four interesting
open questions

1. Is training fair use?

- *Bartz v. Anthropic*: yes
- *Kadrey v. Meta*: yes
- *Thomson Reuters v. ROSS Intelligence*: no

2. Does it matter where the training data came from?

- *Bartz v. Anthropic*:
 - Buying and scanning books is fair use
 - Torrenting scans is not fair use
- *Kadrey v. Meta*:
 - It doesn't matter, since the downstream use that justifies the copying is transformative

3. Does it matter if outputs compete with authors in general?

- *Bartz v. Anthropic*: No, copyright only protects authors against duplicates of their own work
- *Kadrey v. Meta*: Yes, “market dilution” weighs against fair use at the model level, even if all of the outputs are individually transformative enough to be fair use themselves

Are outputs themselves copyrightable?



Zooming out



Anno Octavo

Annæ Reginae.

An Act for the Encouragement of Learning, by Vesting the Copies of Printed Books, in the Authors or Purchasers of such Copies, during the Times therein mentioned.



Whereas Printers, Booksellers, and other Persons have of late frequently taken the Liberty of Printing, Reprinting, and Publishing, or causing to be Printed, Reprinted, and Published Books, and other Writings, without the Consent of the Authors or Proprietors of such Books and Writings, to their very great Detriment, and too often to the Ruin of them and their families: For Preventing therefore such Practices for the future, and for the

Encouragement of Learned Men to Compose and Write useful Books: May it please Your Majesty, that it may be Enacted, and be it Enacted by the Queen's most Excellent Majesty, by and with the Advice and Consent of the Lords Spiritual and Temporal, and Commons in this present Parliament assembled, and by the Authority of the same, That from and after the Tenth Day of April, One thousand seven hundred and ten, the Author of any Book or Books already Printed, who hath not Transferred to any other the Copy or Copies of such Book or Books, Share or Shares thereof, or the Bookeller or Bookellers, Printer or Printers, or other Person or Persons, who hath or have Purchased or acquired the Copy or Copies of any Book or Books, in order to Print or Reprint the same, shall have the sole Right and Liberty of Printing such Book and Books for the Term of One and twenty Years, to Commence from the said Tenth Day of April, and so longer: and that the Author of any Book or Books already Composed and not Printed and Published, or that shall hereafter be Composed, and his Assigns, or Assigns, shall have the sole Liberty of Printing and Reprinting such Book and Books for the Term of four-

?