

 AI Copyright Legal Issues
©

A Tech General Counsel
Turned GPT App Developer
Breaks It Down

The AI Conference

San Francisco
September 10, 2024



Cecilia Ziniti
CEO & Founder, GC AI
3-time GC, ex-Amazon,
ex-MoFo

GC AI 

Pop Quiz - Have you ever

 Trained an AI model?

 Reposted a meme on Twitter?

 Quoted someone in a paper?

 Been inspired by others' art to make your own?

© Been sued for copyright infringement?

Copyright law balances ...

Idea and expression

Old and new

Creatives and Tech

Hmmm.

Copyright in the USA

US Constitution, Article 1, Section 8, Clause 8

—“Congress shall have the power ... to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries.”

Copyright

Exclusive rights granted to authors in original expression

- ✓ To reproduce a work
- ✓ To distribute copies of the work
- ✓ To publicly display or perform
- ✓ To create derivatives

→ You infringe a copyright by doing these without permission

Copyright and AI ...



You

Give me a photorealistic picture of Luigi.



ChatGPT



Here is a photorealistic image of a tall, slender man with a distinctive mustache, wearing a green cap and green shirt, with blue overalls. He is standing in a whimsical, playful fantasy world.



You

show me a plumber brother video game character eating a cranberry tart



ChatGPT



Here is the image of a fictional video game character, resembling a plumber, enjoying a cranberry tart in a colorful, cartoon-style kitchen.



What is fair use?

Permits some uses of copyrighted material
without permission from the © holder.

Four factor test

Fair Use - four factors

1.  /  /  Purpose and character of the use

2.  /  /  Nature of the copyrighted work

3.  /  /  Amount and substantiality of portion used

4.  /  /  Effect of use on market & value of original

NYT v OpenAI

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

THE NEW YORK TIMES COMPANY

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI, INC.,
OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,
OPENAI OPKO LLC, OPENAI GLOBAL LLC,
OAI CORPORATION, LLC, and OPENAI
HOLDINGS, LLC,

Defendants.

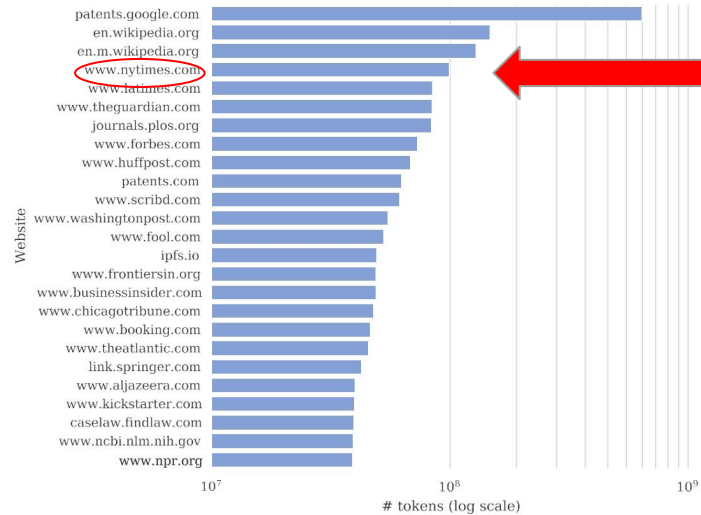
Civil Action No. _____

COMPLAINT

JURY TRIAL DEMANDED

Plaintiff The New York Times Company (“The Times”), by its attorneys Susman Godfrey LLP and Rothwell, Figg, Ernst & Manbeck, P.C., for its complaint against Defendants Microsoft Corporation (“Microsoft”) and OpenAI, Inc., OpenAI LP, OpenAI GP LLC, OpenAI LLC, OpenAI OpCo LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, (collectively “OpenAI” and, with Microsoft, “Defendants”), alleges as follows:

88. The most highly weighted dataset in GPT-3, Common Crawl, is a “copy of the Internet” made available by an eponymous 501(c)(3) organization run by wealthy venture capital investors.²⁰ The domain www.nytimes.com is the most highly represented proprietary source (and the third overall behind only Wikipedia and a database of U.S. patent documents) represented in a filtered English-language subset of a 2019 snapshot of Common Crawl, accounting for 100 million tokens (basic units of text):²¹



Output from GPT-4:

exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed.

Their actions turned one of the best-known symbols of New York — its yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat.

“Nobody wanted to upset the industry,” said David Klahr, who from 2007 to 2016 held several management posts at the Taxi and Limousine Commission, the city agency that oversees medallions. “Nobody wanted to kill the golden goose.”

New York City in particular failed the taxi industry, The Times found. Two former mayors, Rudolph W. Giuliani and Michael R. Bloomberg, placed political allies inside the Taxi and Limousine Commission and directed it to sell medallions to help them balance budgets and fund key initiatives.

During that period, much like in the mortgage lending crisis, a group of industry leaders enriched themselves by artificially inflating medallion prices. They encouraged medallion buyers to borrow as much as possible and ensnared them in interest-only loans and other one-sided deals that often required borrowers to pay hefty fees, forfeit their legal rights and give up most of their monthly incomes.

When the market collapsed, the government largely abandoned the drivers who bore the brunt of the crisis. Officials did not bail out borrowers or persuade banks to soften loan

Actual text from NYTimes:

exempted it from regulations, subsidized its operations and promoted its practices, records and interviews showed.

Their actions turned one of the best-known symbols of New York — its signature yellow cabs — into a financial trap for thousands of immigrant drivers. More than 950 have filed for bankruptcy, according to a Times analysis of court records, and many more struggle to stay afloat.

“Nobody wanted to upset the industry,” said David Klahr, who from 2007 to 2016 held several management posts at the Taxi and Limousine Commission, the city agency that oversees cabs. “Nobody wanted to kill the golden goose.”

New York City in particular failed the taxi industry, The Times found. Two former mayors, Rudolph W. Giuliani and Michael R. Bloomberg, placed political allies inside the Taxi and Limousine Commission and directed it to sell medallions to help them balance budgets and fund priorities. Mayor Bill de Blasio continued the policies.

Under Mr. Bloomberg and Mr. de Blasio, the city made more than \$855 million by selling taxi medallions and collecting taxes on private sales, according to the city.

But during that period, much like in the mortgage lending crisis, a group of industry leaders enriched themselves by artificially inflating medallion prices. They encouraged medallion buyers to borrow as much as possible and ensnared them in interest-only loans and other one-sided deals that often required them to pay hefty fees, forfeit their legal rights and give up most of their monthly incomes.

OpenAI has good arguments



While we disagree with the claims in The New York Times lawsuit, we view it as an opportunity to clarify our business, our intent, and how we build our technology. Our position can be summed up in these four points, which we flesh out below:

1. We collaborate with news organizations and are creating new opportunities
2. Training is fair use, but we provide an opt-out because it's the right thing to do
3. "Regurgitation" is a rare bug that we are working to drive to zero
4. The New York Times is not telling the full story



Suno and Udio Cases

Case 1:24-cv-04777-AKH Document 26 Filed 08/01/24 Page 1 of 40

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

UMG RECORDINGS, INC., CAPITOL
RECORDS, LLC, SONY MUSIC
ENTERTAINMENT, ARISTA MUSIC, ARISTA
RECORDS LLC, ATLANTIC RECORDING
CORPORATION, RHINO ENTERTAINMENT
COMPANY, WARNER MUSIC INC., WARNER
MUSIC INTERNATIONAL SERVICES
LIMITED, WARNER RECORDS INC.,
WARNER RECORDS LLC, and WARNER
RECORDS/SIRE VENTURES LLC,

Plaintiffs,

v.

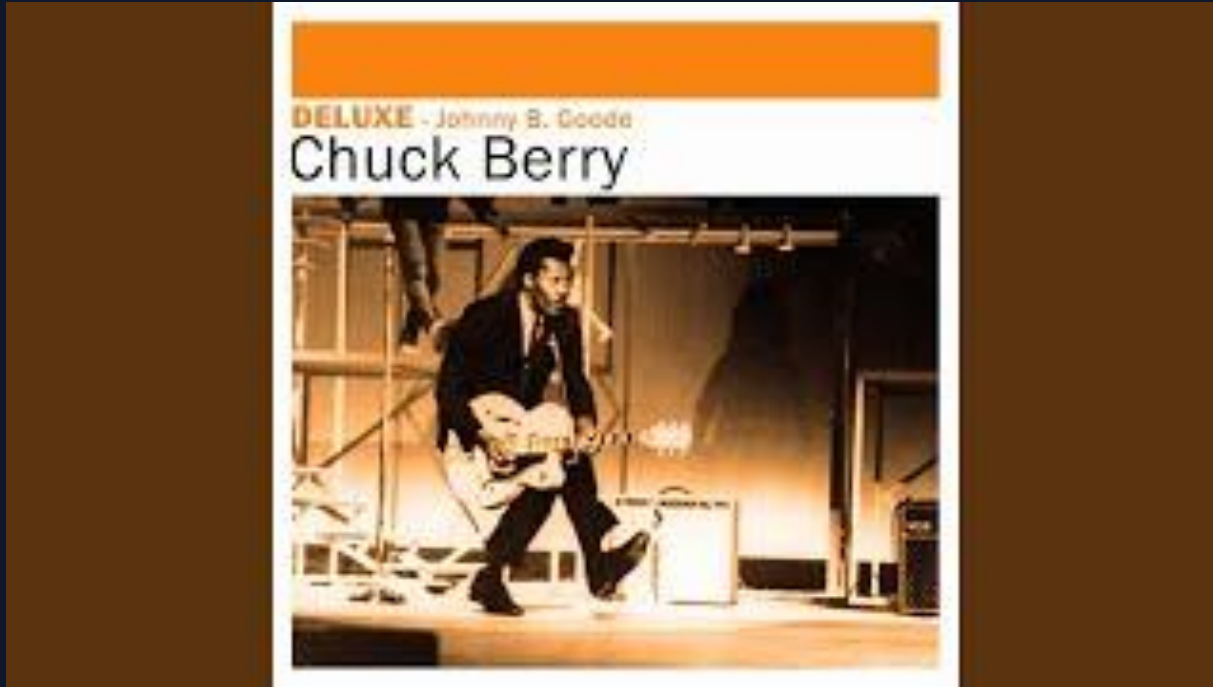
UNCHARTED LABS, INC., d/b/a Udio.com, and
JOHN DOES 1-10,

Defendant.

Case No. 1:24-cv-04777-AKH

Hon. Alvin K. Hellerstein

Original Johnny B Goode Versus Suno Version



Original Version

Original Johnny B Goode Versus Suno Version




SUNO



- Home
- Create
- Library
- Explore
- Search

Subscribe

What's New? **20**






Help



+  

Deep down in Louisiana close to New Orle

1950s rock and roll, 12 bar blues, rhythm & blues, rockabilly, energetic male vocalist

 GalvanizingAudiotogra... [Follow](#)  **1.3K**  **2**  

May 28, 2024 at 5:53 PM [v3.5](#)

[Verse]
Deep down in Louisiana close to New Orleans
Way back up in the woods among the evergreens
There stood a log cabin made of earth and wood
Where lived a country boy named Johnny B. Goode
Who never ever learned to read or write so well
But he could play a guitar just like a-ringin' a bell

[Chorus]
Go, go,
Go Johnny, go, go
Go Johnny, go, go
Go Johnny, go, go
Go Johnny, go, go
Johnny B. Goode

Suno Version

Suno and Udio Cases

copyrights. v.v.v.v.

Fair Use Defense From the Suno Answer:

Plaintiffs explicitly disavow any contention that any output ever generated by Suno has ever infringed any right that they own. *Id.* While the Complaint includes a variety of examples of outputs that allegedly resemble certain pre-existing songs, *id.* ¶¶ 51–67, it goes out of its way to say that the Complaint is not alleging that those outputs constitute actionable copyright infringement, *id.* ¶ 50 (“Plaintiffs are not . . . alleging that these outputs themselves infringe the Copyrighted Recordings”).

So can you use train an AI model?

Yes, with some pro tips.



Open internet generally ok



Know your data



Document what you do



Lawyers at the get-go



Pass the smell test always



If your biz model is sneaky, be ready to defend



Once you launch, story and comms matter

What about AI apps?

✓ Using big foundation models is OK. Risk low as plaintiffs likely to sue the source.



Have a takedown process that is FAST



Test for glaring regurgitation



Solid terms of use putting onus on your users



Solid comms when you launch

Fun Cases (time permitting)

✔ Betamax Case - time shifting: substantial non-infringing use for VCR



✓ Pretty Woman tune & 1 verse as rap and & raunchy - fair



Campbell v Acuff-Rose (1994)

✓ *Sony v Connectix* -
copying to make an emulator
is fair use

9th Cir. 2000

Connectix Corporation



Connectix

Industry	Computer software and hardware
Founded	October 1988
Defunct	August 2003
Fate	Sold QuickCam to Logitech, sold Virtual Game Station to Sony, sold Virtual PC to Microsoft, discontinued other products

Connectix went bust anyway

✓ Screenshotting
video game for
emulator's ad - fair

*Sony Computer Entertainment America
v. Bleem, 214 F.3d 1022 (9th Cir. 2000)*

PLAY HUNDREDS OF PLAYSTATION GAMES ON YOUR PC! WIN95/98 CD-ROM

bleem!

**GET NEXT-GENERATION GAMEPLAY TODAY!
HIGHER RESOLUTION • MORE DETAIL • RICHER COLOR**

**PLAY PLAYSTATION GAMES ON YOUR PC
- LIKE YOU'VE NEVER SEEN THEM BEFORE!**

PLAY PLAYSTATION GAMES ON YOUR PC! Bleem! from BLEM! LLC is great PlayStation games on your PC, with no extra hardware required. Its Software Match Agent™ delivers great compatibility and PlayStation-level graphics. Limited only by the speed of your PC, all the games you play - this time, you get great results with many titles!

NEXT-GENERATION ENHANCED GRAPHICS. And if you're a power 3D accelerator card fiend - 16.7 million colors and more, the next! Many compatible games are also "3D Enhanced" - revealed in standard 1024x768 PlayStation resolution, you can play at 848x480, 1024x768, and even higher - for unbelievable clarity and detail. With Bleem! and a power 3D accelerator, even older PCs can produce the HDing game quality!

TAKE YOUR GAMES WITH YOU! Finally a cure for home game trip! Bleem! means you can take your favorite games anywhere you can bring your laptop. And since Bleem! is 100% software, you can easily "test the waters" and enjoy your games at the office, at school, at grandma's, there's a PC!

FREE UPDATES! Bleem! constantly works to make better, with added support for new releases and more titles. And you can get the latest updates and patches. Don't miss a thing! Get the latest version of an update, see FREE! from www.bleem.com

PERFECT FOR EVERY GAME. If you already own a PlayStation, Bleem! gives you next-generation gameplay from the library you already own - ALL! Finally, even the best for TV line. And if you're a PC gamer, now you can do just that! Hundreds of great PlayStation games, right on your PC. For better today!

WHAT WORKS WITH BLEEM!? Bleem! plays over 100 of the most popular PlayStation titles. For a complete list, visit our website. Check out the full feature overview!

OS SUPPORT: Windows 95, 98, and Mac OS 9.0+.

CD-ROM SUPPORT: Yes, you need a CD-ROM drive to play the games on CD-ROM.

DRIVE LETTERS: Windows 95 & 98 require a C: drive. Mac OS 9.0+ requires a drive letter.

GAME TITLE: All game titles must be entered exactly as they appear on the box. Some titles may have alternate names.

YOU WILL NEED: Windows 95 or 98, a CD-ROM drive, a power 3D accelerator card, and a monitor.

BLEEM! IS NOT: A game, a CD-ROM, or a software patch. It is a software application that runs on your PC.

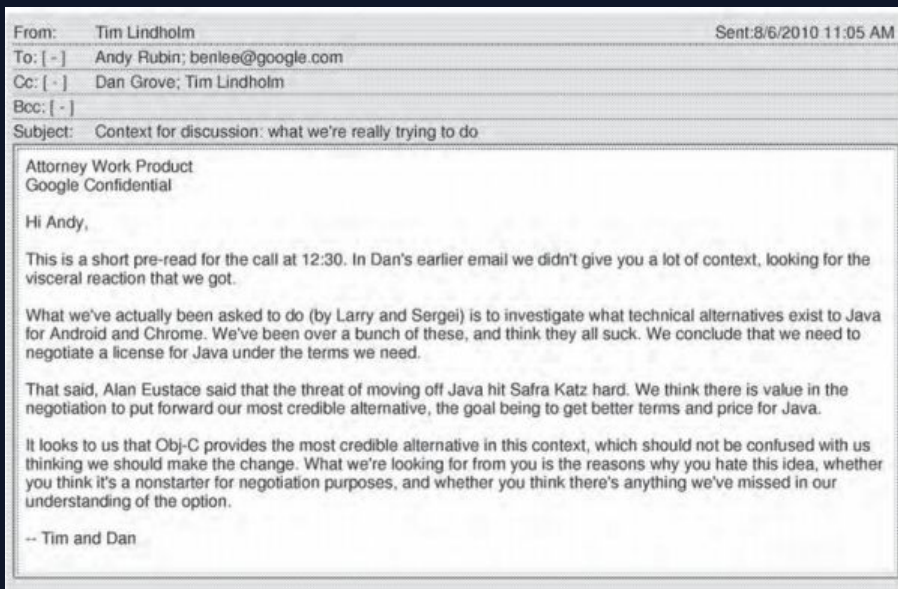
AND MORE!

TURN THIS: WITH BLEEM! AND YOUR 3D GRAPHICS CARD, YOU CAN: **INTO THIS!** AND MORE!

©2000 Sony Computer Entertainment Inc. All rights reserved. PlayStation is a registered trademark of Sony Computer Entertainment Inc. Bleem! and Bleem! logo are trademarks of BLEM! LLC.



Google copying of Java code for interoperability

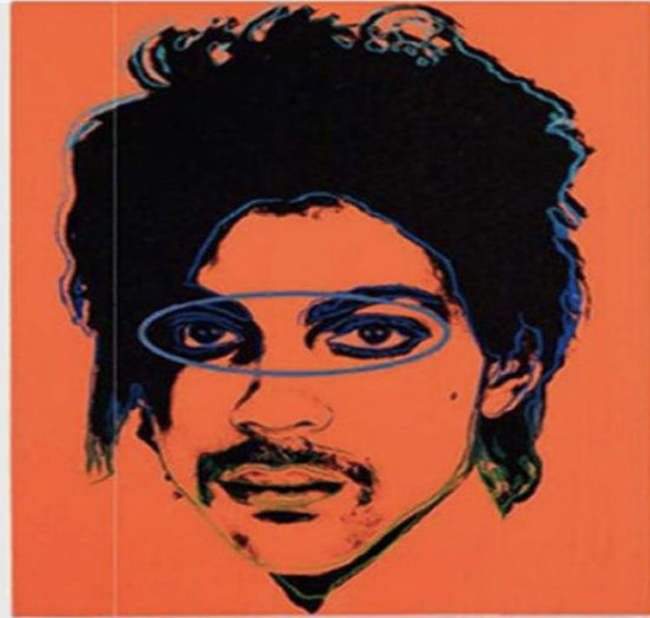


✗ Copying doll's exact details so they're like twins



^^ Doll on the right ruled to infringe the doll on the left - it's not the idea of a farting doll that's copyrighted, it's the expression of it. Here - sitting, blue pants, dark hair, weird smile.

✗ Colorizing and Reprinting a Prince picture - not transformative (*Warhol*)







✘ “Rusty Krab” restaurant - after “Krusty Krab” is in many episodes of copyrighted Sponge Bob series



Appendix

Fair Use today - Section 107 of the Copyright Act

-  Purpose and character of the use
→ Key for the AI cases
-  Nature of the copyrighted work
-  Amount and substantiality of portion used
-  Effect of use on market & value of original

 /  /  Purpose and character of the use
→ Key in NYT v. OAI

Fair Use Cases

✓ VTR - OK because time-shifting is real use (*Betamax*)

✓ Pretty Woman tune, verse, but risqué and raunchy
(*Campbell*)

✓ Copying so you can make a console emulator (*Connectix*)

✓ Screenshots of original game in emulator's ad (*Bleem*)

✗ Antennas for OTA broadcast → internet (*Aereo*)

✓ Google copying Java code for interop (0.4% of whole)

✗ Colorizing photo licensed “for one time use” (*Warhol*)

✗ “Rusty Krab” restaurant infringes Sponge Bob (*Viacom v
Pixar*)

Weighing the Four Factors

- Not a simple sum
- Factors aren't binary
- Relative weight varies depending on circumstances.
- Factor 1 and Factor 4 are generally thought to be the most important.
 - But not always true; other factors can be determinative
 - Harper & Row



We support journalism, partner with news organizations, and believe The New York Times lawsuit is without merit.



1. We collaborate with news organizations and are creating new opportunities
2. Training is fair use, but we provide an opt-out because it's the right thing to do
3. "Regurgitation" is a rare bug that we are working to drive to zero
4. The New York Times is not telling the full story

OpenAI and Journalism

Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.



Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.

Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.

Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.

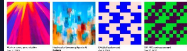
Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.

Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.

Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.

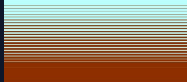
Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.

Related research



Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.

Journalism is a profession that has been around for centuries, and it has always been a profession that has been in the vanguard of change.





While we disagree with the claims in The New York Times lawsuit, we view it as an opportunity to clarify our business, our intent, and how we build our technology. Our position can be summed up in these four points, which we flesh out below:

1. We collaborate with news organizations and are creating new opportunities
2. Training is fair use, but we provide an opt-out because it's the right thing to do
3. "Regurgitation" is a rare bug that we are working to drive to zero
4. The New York Times is not telling the full story



1. We collaborate with news organizations and are creating new opportunities

We work hard in our technology design process to support news organizations.

We've met with dozens, as well as leading industry organizations like the News/Media Alliance, to explore opportunities, discuss their concerns, and provide solutions. We aim to learn, educate, listen to feedback, and adapt.

Our goals are to support a healthy news ecosystem, be a good partner, and create mutually beneficial opportunities. With this in mind, we have pursued

2. Training is fair use, but we provide an opt-out because it's the right thing to do

Training AI models using publicly available internet materials is fair use, as supported by long-standing and widely accepted precedents. We view this principle as fair to creators, necessary for innovators, and critical for US competitiveness.

The principle that training AI models is permitted as a fair use is supported by a wide range of academics, library associations, civil society groups, startups, leading US companies, creators, authors, and others that recently submitted comments to the US Copyright Office. Other regions and countries, including the European Union, Japan, range of academics, library associations, civil society groups, startups, leading US companies, creators, authors, and others that recently submitted comments to the US Copyright Office. Other regions and countries, including the European Union, Japan, Singapore, and Israel also have laws that permit training models on copyrighted content—an advantage for AI innovation, advancement, and investment.

That being said, legal right is less important to us than being good citizens. We have led the AI industry in providing a simple opt-out process for publishers (which The New York Times adopted in August 2023) to prevent our tools from accessing their sites.



**AUTHORS
ALLIANCE**

authorsalliance.org
510.480.8302

2705 Webster St. #5805
Berkeley, CA 94705

Founding Members

Jasmine Abdel-khalik

UMKC School of Law

Alessandro Acquisti

Carnegie Mellon University

Prue Adler

Association of Research
Libraries

Julie Ahrens

Stanford Law School,
Center for Internet &
Society

Ivy Anderson

California Digital Library

Jonas Anderson

American University
Washington College of Law

Patricia Aufderheide

American University

A. Michael Froomkin

University of Miami School
of Law

Laura Gasaway

University of North
Carolina School of Law

Steve Gass

MIT

Andy Gass

Paul Geller

Ilana Gershon

Indiana University

Shubha Ghosh

Rebecca Giblin

Faculty of Law, Monash
University, Australia

James Gibson

Jack Lerner

Karen Levy

Cornell University

Roger Levy

MIT

Yvette Joy Liebesman

St. Louis University School
of Law

Jacqueline Lipton

Authography LLC

Jessica Litman

University of Michigan

Lydia Loren

Lewis & Clark Law School

Brian Love

Santa Clara University
School of Law

Julie Samuels

Tech:NYC

Sharon Sandeen

Mitchell Hamline School of
Law

Joshua Sarnoff

DePaul University College
of Law

AnnaLee Saxenian

UC Berkeley

Rich Schneider

UC San Francisco

Jason Schultz

NYU School of Law

Michael Scott

Southwestern Law School

Jeffrey Selbin

UC Berkeley Law

mutually beneficial opportunities. With this in mind, we have pursued partnerships with news organizations to achieve these objectives:

1. Deploy our products to benefit and support reporters and editors, by assisting with time-consuming tasks like analyzing voluminous public records and translating stories.
2. Teach our AI models about the world by training on additional historical, non-publicly available content.
3. **Display real-time content with attribution in ChatGPT, providing new ways for news publishers to connect with readers.**

Our early partnerships with the Associated Press, Axel Springer, American Journalism Project and NYU offer a glimpse into our approach.

3. **“Regurgitation” is a rare bug** that we are working to drive to zero

Our models were designed and trained to learn concepts in order to apply them to new problems.

Memorization is a rare failure of the learning process that we are continually making progress on, but it's more common when particular content appears more than once in training data, like if pieces of it appear on lots of different public websites. So we have measures in place to limit inadvertent memorization and prevent regurgitation in model outputs. We also expect our users to act responsibly; intentionally manipulating our models to regurgitate is not an appropriate use of our technology and is against our terms of use.

Just as humans obtain a broad education to learn how to solve new problems, we want our AI models to observe the range of the world's information, including from every language, culture, and industry. Because models learn from the enormous aggregate of human knowledge, any one sector—including news—is a tiny slice of overall training data, and any single data source—including The New York Times—is not significant for the model's intended learning.

Submission of Ad Hoc Group of Artists Using Generative AI¹

We, the undersigned professional artists who use generative AI tools, welcome this opportunity

[Heather Timm](#)
[Nettrice Gaskins](#)
[T. K. Yonge](#), [Grey Owl Audio](#)
[wondermundo](#)
[Henry Daubrez](#)
[dadabots](#)
[Jrdscct](#)
[Tim Boucher](#)
[makeitrad](#)
[FARRAH CARBONELL](#)
Michael Foxington
[Pat Tremblay](#)
[F Disa](#)
[Jordan McRae](#)
[espinosa](#)
[Mikki Sharp](#)
AD Campbell
OxCrumpet
Pixlosopher
D O'Donnell
[Gushousekai195](#)
[Niki Selken](#)

[Gushousekai195](#)
[Niki Selken](#)
Tim Newman
[Johanna Wilder](#)
[Roberta \(Robin\) Sullivan](#)
[KEZIAI](#)
[Vasiawow](#)
[MsJoKnows](#)
[Nell Fallcard](#)
[Purz](#)
[WEBmadman](#)
[Web Horizons](#)
[Mario Klingemann](#)
Nucksen
JosephC
[Elizabeth Ann West](#)
[@AlandDesign on X](#)

[ME Malainig](#)
[Robert Paul AKA MxVoid](#)
[TheDigitalCov](#)
[Manu Williams](#)
Antoni Albalat
[Tim Simpson Jr.](#)
[Celestino Espinoza L.](#)
[Artful Insanity](#)
[NIFTey](#)
[Gushousekai195](#)
[Esterina Valimikova](#)
[Souki Mansoor](#)
[SBodan](#)
[Anil](#)
[Brian L. Frye](#)
[Niccolo Casas](#)
[Rebel Without Applause](#)
[4everKurious](#)
Thom Hofnarreth
[Aporium](#)
[Meta is Done by Joseph Buchanan and Jennifer Ruff](#)
neville glover
Snowy79
Nalani
[AtikoAE](#)
[▲▲▲](#)
[Sivasense](#)
[PJ Way](#)

[T. K. Yonge](#)
[ANTHONY GALIGO](#)
[Ai Machina](#)
[XTincT](#)
[Carlos28355](#)
[dr.mindset](#)
Nima Leo
[oh_x_d](#)
[SouthernFried](#)
[Miriam Dunn](#)
Fatima
[@MagicallyViral](#)
[Brooke Ganster](#)
[Tinas_art_beat](#)
[Zen Materialist](#)

Fair Use

- Notoriously difficult to predict
- Breakdown of potentially fair uses:
 - Some clearly unfair uses
 - Fewer are clearly fair
 - The majority fall within a vast gray area
- What's the point?
 - Recognize the clear cases
 - Recognize when it makes sense to call a lawyer

The Four Fair Use Factors

- (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- (4) the effect of the use upon the potential market for or value of the copyrighted work.
 - These 4 factors are not exclusive; courts can and do consider other factors
 - Public interest

Factor 1

- (1) the **purpose and character** of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
 - Is the alleged infringer making for-profit or non-profit use?
 - Is the alleged infringer making a transformative use?
 - A transformative use is one that adds new meaning or purpose to the original work
 - > Parody, Search tools

Factor 2

- (2) the **nature** of the copyrighted work;
 - Was the work published or unpublished?
 - Publication favors fair use
 - Was the work primarily factual or fictional?
 - Spectrum between factual and fictional
 - The more creative the work, the less likely the use will be found fair

Factor 3

- (3) the **amount and substantiality** of the portion used in relation to the copyrighted work as a whole;
 - The more the alleged infringer uses, the less likely the use will be fair
 - In some cases, using the entire work is fair
 - In others, using a small fraction of a work is unfair, particularly if the portion used is “the heart of the original work”

Factor 4

- (4) the **effect** of the use **upon the potential market** for or value of the copyrighted work.
 - Does the allegedly infringing work serve as a market substitute for the original work or deprive the copyright holder of revenue?
 - If a parody harms sales as a result of its criticism, this cannot weigh against fair use
 - Look not only to current markets, but future potential markets as well

>Problem of circularity

Scraping

- US law:
 - Web scraping is generally legal if the information is publicly accessible, but it can be restricted by website terms and conditions.

- Three key cases
 - hiQ Labs v. LinkedIn (public data scraping is legal)
 - Facebook v. Power Ventures (cease and desist revokes authorization)
 - Craigslist v. 3Taps (website terms can limit scraping)

Pro tips for AI developers

- IP strategy
 - Data provenance
 - Keep good records
 - Document your reasoning
 - Does it pass the smell test
 - Hire good lawyers!

- Training models legally: Focus on transparency, obtain necessary permissions, and document your processes to demonstrate compliance.

So can you train a model? You can

- Know your data
-
-
- Legal considerations: Developers must navigate copyright law when training AI models, considering the source and nature of the data used.
- Best practices: Use publicly available data, seek licenses where necessary, and consider fair use implications.