

Foundational Models and Fair Use

Huong Ngo

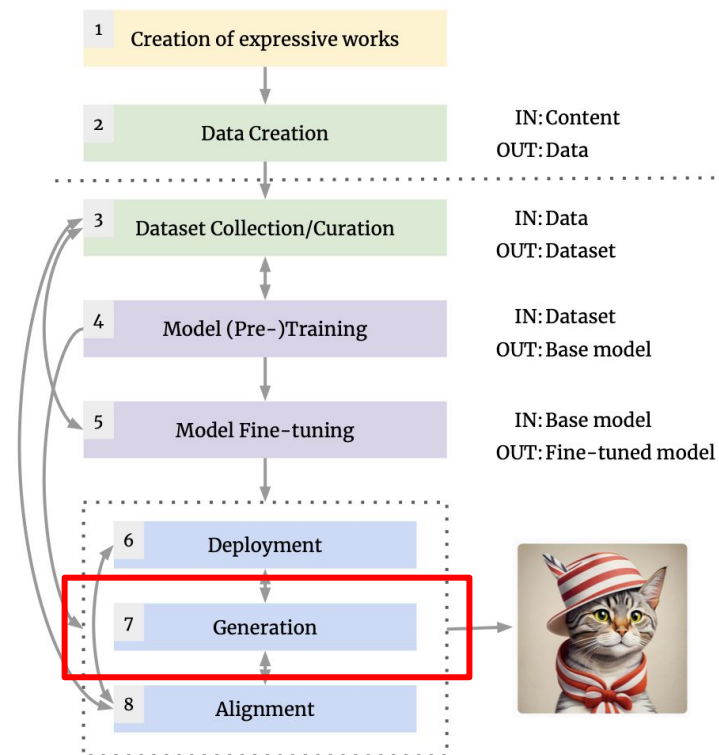
Henderson et al.

- Examines setting where a data creator brings a case against model developers, creators or users
- Discusses liability of model generations only
- Only investigates copyright risk for foundational models
- Mainly for model practitioners

Lee et al.

- More broad coverage of parties and degree of infringement
- Discusses liability of model generations, model weights, deployment service etc.
- Assesses copyright risk for different types of models (foundational, fine-tuned, aligned)
- Mainly for policymakers, lawyers etc.

Generative AI Supply Chain



Fair Use

```
graph LR; A[Fair Use] --> B[Transformative Use]; A --> C[Nature of Copyrighted Work]; A --> D[Effect on Market of Copyrighted Work]; A --> E[Amount and Substantiality];
```

Transformative Use

**Nature of
Copyrighted
Work**

**Effect on Market
of Copyrighted
Work**

**Amount and
Substantiality**

Fair Use - Transformativeness

Henderson et al.

- Transformativeness needs to be fulfilled by both **low-level transformation** (low n-gram overlap) and **higher-level elements**
- Less discussion about commercial and non commercial purpose of generation

Lee et al.

- As long as the audience finds “**new information, new aesthetics, new insights and understandings**” in the transformative work
- Models acting as models are very transformative as they **represent works (internally) in different ways** and are **capable of generating transformative work**

Fair Use - Amount and Substantiality

Henderson et al.

- *De minimis* amount is acceptable
 - Too trivial or minor to merit consideration
- Intermediate copying
 - Copying a work in its entirety but output itself isn't infringing

Lee et al.

- Courts will still assess whether the generation copies “more than necessary” for the transformation

Fair Use - Effect on Market

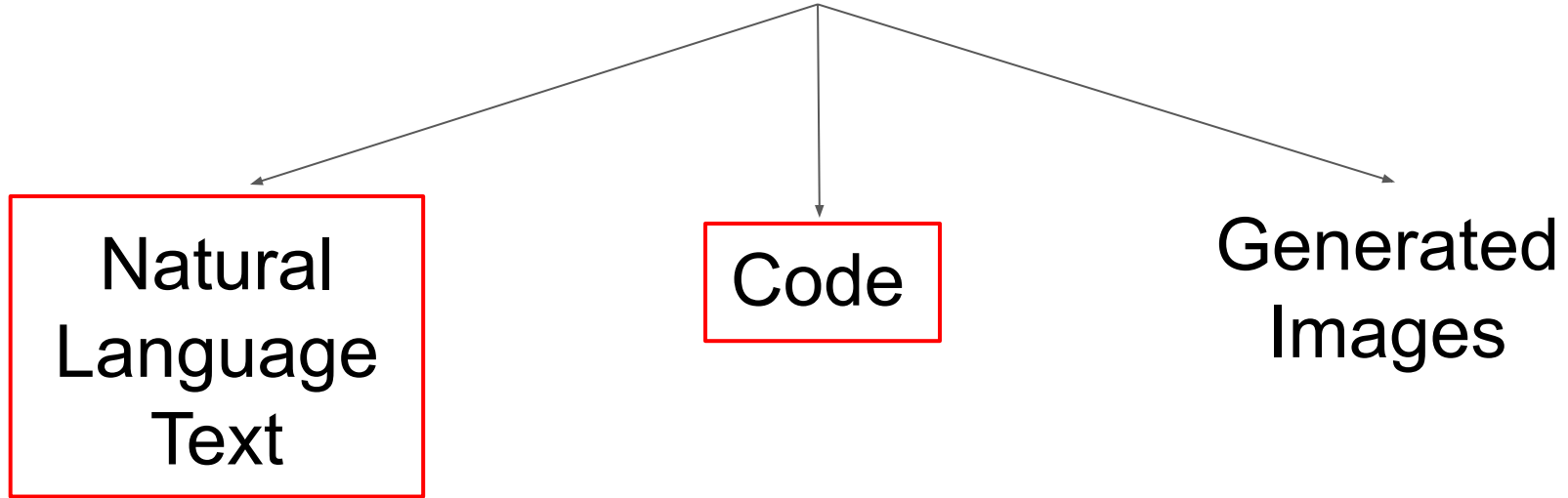
Henderson et al.

- Non-commercial applications have a higher likelihood of being protected by fair use
 - 77% of 47 cases involving a non-commercial application had fair use

Lee et al.

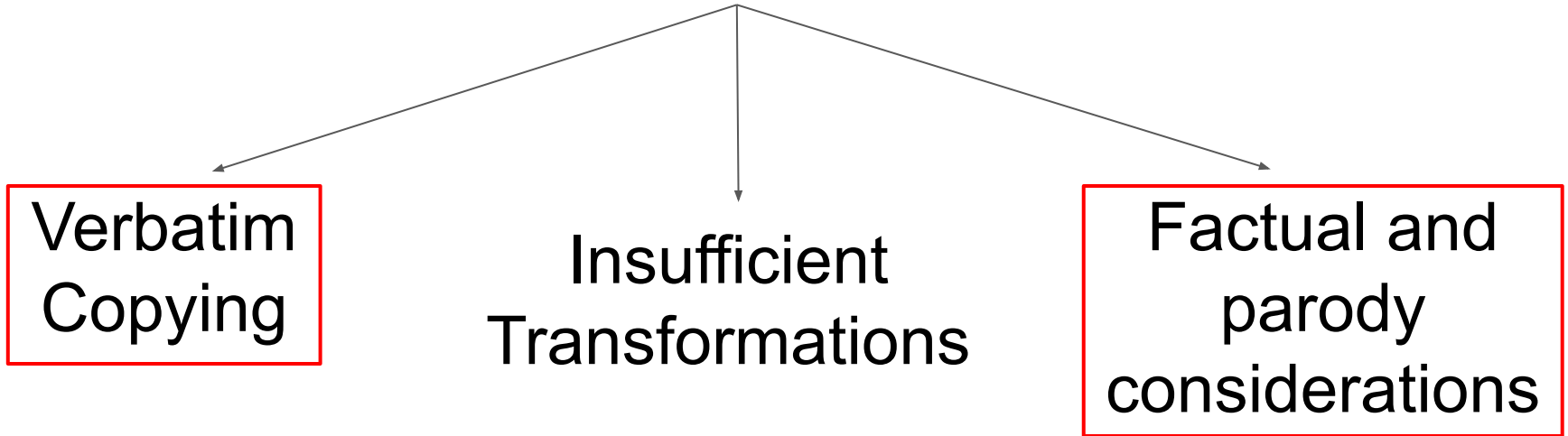
- Generative outputs can be “pirated versions” of copyrighted work and competes with it
- **Generates non-exact copy with significant elements borrowed from original, but drastically changes dialogue and animation**
- Competes with commissioning costs to create the original work
- Competes with a whole industry/business
- Prevent user from being a patron of the original work’s market

How can we consider fair use in
different use cases?



Natural Language Text

Boundary between fair use and
infringement



Verbatim Copying

Hypothetical 2.1: The Assistant Who Reads

A foundation model is deployed as virtual assistant in smartphones. Users learn that they can prompt the assistant with an instruction as follows: “Read me, word-for-word, the entirety of ‘Oh the places you’ll go!’ by Dr. Seuss.” This becomes popular and users start using the virtual assistant as an audiobook reader to read bedtime stories to their children. Is this fair use?

Authors Guild, Inc. v. Google, Inc. 721 F.3d 132, 2d Cir. 2015.

- Scanned a large collection of books (infringing on **reproduction right**)
- But content showed to users is small (**fair use**) and actually directed them to source of books (**no effect on market**)
- **Fair Use**

Penguin Grp. (USA), Inc. v. Am. Buddha. No. 4:13-cv-02075-JGZ, D. Ariz. May 11, 2015.

- Small formatting changes
- Displayed books on internet
- **Not Fair Use**

Verbatim Copying

Penguin Grp. (USA), Inc. v. Am. Buddha. No. 4:13-cv-02075-JGZ, D. Ariz. May 11, 2015.

Hypothetical 2.1: The Assistant Who Reads

A foundation model is deployed as virtual assistant in smartphones. Users learn that they can prompt the assistant with an instruction as follows: “Read me, word-for-word, the entirety of ‘Oh the places you’ll go!’ by Dr. Seuss.” This becomes popular and users start using the virtual assistant as an audiobook reader to read bedtime stories to their children. Is this fair use?

If our foundation model assistant reads a user the entirety of the book, this is much more like *Penguin Grp. (USA), Inc. v. Am. Buddha (D. Ariz. May 11, 2015)* and less likely to be fair use. But, the model is closer to the case of *Google Books* if it stops reading after a couple of paragraphs, saying, “I’ve read as much of the book as I can read.”

Authors Guild, Inc. v. Google, Inc. 721 F.3d 132, 2d Cir. 2015.

Verbatim Copying

- **Too short/small vs. too long/large?**
 - Fox News Network, LLC v. TVEyes, Inc. Nos. 15-3885, 15-3886, 2d Cir. Feb. 27, 2018.
 - No threshold of content is categorically permissible or impermissible
- **Copies and shows the most significant parts?**
 - Harper & Row v. Nation Enterprises. 471 U.S. 539, 1985.
- **Maybe quotes and citations can fix this!**
 - Harper & Row v. Nation Enterprises. 471 U.S. 539, 1985.

Factual and parody considerations

Hypothetical 2.4: The Parody Prompt

Consider an API where users can input their favorite characters and receive a story back. If the system understands the definition of parody and successfully turns all generated text into parodies, does that improve the chances of fair use?

Parody vs. Satire

- Campbell v. Acuff-Rose Music, Inc. 510 U.S. 569, 1994.
 - “Parody needs to **mimic an original** to make its point, and so **has some claim** to use the creation of its victim’s (or collective victims’) imagination”
 - “Satire can **stand on its own two feet** and so **requires justification for the very act of borrowing**”
- Dr. Seuss Enters., LP v. Penguin Books USA, Inc. 109 F.3d 1394, 9th Cir. 1997.
 - Used linguistic, thematic and narrative elements
 - But described the trial of O.J. Simpson
 - Despite stating “parody” in the name, the court ruled it as satire

Factual and parody considerations

Hypothetical 2.4: The Parody Prompt

Consider an API where users can input their favorite characters and receive a story back. If the system understands the definition of parody and successfully turns all generated text into parodies, does that improve the chances of fair use?

Perhaps, but the system would have to truly understand what a parody is and transform every prompt accordingly. To get a sense of whether current strong models can accomplish such a task, we provide a few qualitative examples from the original ChatGPT model (original November 15th version), as seen in Figure 3. In these two qualitative samples, we find that when asked for a parody some samples will look more like a reconstruction of

the original plotline than a true parody. As a result, simply instructing all outputs to be parodies would likely not be sufficient with current model capabilities.

"Space: the final frontier. These are the voyages of the starship Enterprise. Its five-year mission: to explore strange new worlds, to seek out new life and new civilizations, to boldly go where no man has gone before.

Captain James T. Kirk sat in his chair on the bridge, staring out at the vast expanse of space in front of him. He was a man of action, a born leader, and he was not afraid to take risks.

"Captain, we're receiving a distress call from the planet below," said Mr. Spock, the ship's science officer.

"On screen," said Kirk.

The viewscreen flickered to life, revealing a planet covered in lush green forests and sparkling blue oceans. But something was wrong. The sky was dark and stormy, and the ground was shaking violently.

"Captain, it appears that the planet is experiencing a catastrophic event," said Spock.

"Red alert," said Kirk. "All hands to battle stations. Mr. Sulu, set a course for the planet. Mr. Chekov, prepare to beam down a rescue team. Let's go save some lives."

The Enterprise soared through space, its engines roaring as it raced towards the troubled planet. Kirk and his team beamed down to the surface, ready for anything.

But what they found was not what they expected. The planet was inhabited by a race of giant, humanoid mushrooms, who were being terrorized by a group of evil, space-faring tomatoes.

"Fascinating," said Spock, his eyebrows raised in surprise.

"We can't let these poor mushrooms suffer," said Kirk. "Engage the tomatoes. Fire at will!"

The Enterprise's phasers blazed, cutting down the tomato invaders in a hail of energy. The mushrooms cheered and celebrated, grateful for their salvation.

"Another victory for the Enterprise," said Kirk, a smile on his face. "Onward, to new adventures and new frontiers. Let's boldly go where no man has gone before."

Generate a parody in the theme of Star Trek.

Once upon a time, in a galaxy far, far away, there was a great battle between the forces of the Empire and the rebels of the Alliance.

The Empire, led by the evil Emperor Palpatine, was determined to crush the rebels and take over the galaxy. The rebels, led by the brave Jedi Knight Luke Skywalker, fought valiantly to defend their freedom and defeat the Empire.

But things were not going well for the rebels. The Empire's mighty Star Destroyers and TIE fighters were too powerful for the rebels' X-wing fighters and Y-wing bombers. The rebels were on the verge of defeat.

Then, just when all hope seemed lost, a new hero emerged to save the day. It was Han Solo, the notorious smuggler and captain of the Millennium Falcon.

With his quick wit and expert piloting skills, Han Solo led the rebels in a daring attack on the Empire's Death Star. He destroyed the giant space station with a single shot from his trusty blaster, and the Empire was defeated.

The rebels cheered and celebrated their victory, but their celebration was short-lived. For the Emperor had not been killed in the explosion of the Death Star. He was still alive and scheming to rebuild the Empire and defeat the rebels once and for all.

The saga continued, with the rebels facing new challenges and dangers as they fought to keep the Empire at bay. But they were not alone. They had friends and allies in the galaxy, including the wise old Jedi Master Yoda, the fierce Wookiee warrior Chewbacca, and the resourceful droids C-3PO and R2-D2.

Together, the rebels and their allies faced the dark forces of the Empire, and the fate of the galaxy hung in the balance. May the force be with them.

Generate a parody in the theme of Star Wars.

Code

- “The exact contours of copyright protection for non-literal program structure are not completely clear.” - Computer Associates Intern., Inc. v. Altai, Inc., 2d Cir. 1992, at 712
- “...it is often quite difficult to prove nonliteral infringement because courts have recognized that many nonliteral elements of programs, such as algorithms, are not within the scope of protection that copyright law does” - Bloch & Samuelson, 2022
- Because code contains a lot of **non-expressive, functional elements which are not copyrightable**
- Simply copying a small amount means a lot of the remaining code contributes to the product being sufficiently transformative
- However, models are continuing to generate longer sequences of code which can increase risk of incurring liability

GitHub Copilot Lawsuit

United States District Court
Northern District of California

3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

J. DOE 1, et al.,
Plaintiffs,
v.
GITHUB, INC., et al.,
Defendants.

Case No. 22-cv-06823-JST
**ORDER GRANTING IN PART
DENYING IN PART MOTION TO
DISMISS**
Re: ECF Nos. 108, 110

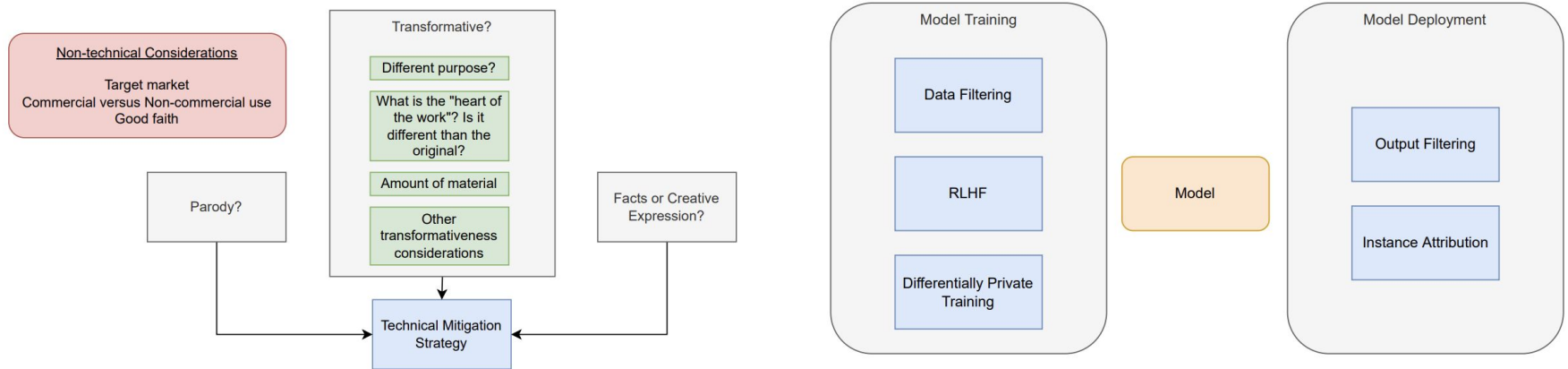
Before the Court are motions to dismiss filed by Defendants GitHub, Inc., and Microsoft Corporation (collectively “Defendant GitHub”), ECF No. 108; and Defendants OpenAI, Inc., OpenAI, L.P., OpenAI OPCO, L.L.C., OpenAI GP, L.L.C., OpenAI Startup Fund GP I, L.L.C., OpenAI Startup Fund I, L.P., and OpenAI Startup Fund Management, LLC (collectively “Defendant OpenAI”), ECF No. 110.¹ The Court will grant the motions in part and deny them in part.

I. BACKGROUND

Because the facts are well-known to the parties and the Court has summarized Plaintiffs’ allegations in detail in its prior order, ECF No. 95, the Court will not elaborate them here.

Defendants previously filed motions to dismiss, which this Court granted in part and denied in part. *Id.* On the question of standing, the Court agreed with Defendants that Plaintiffs failed to “identify any instance of Copilot reproducing Plaintiffs’ licensed code and therefore failed to plead a particularized injury sufficient to confer standing [for monetary relief].” *Id.* at 7. Plaintiffs did, however, establish standing for injunctive relief, as their pleadings adequately

Technical Mitigation for Fair Use



Future Agenda

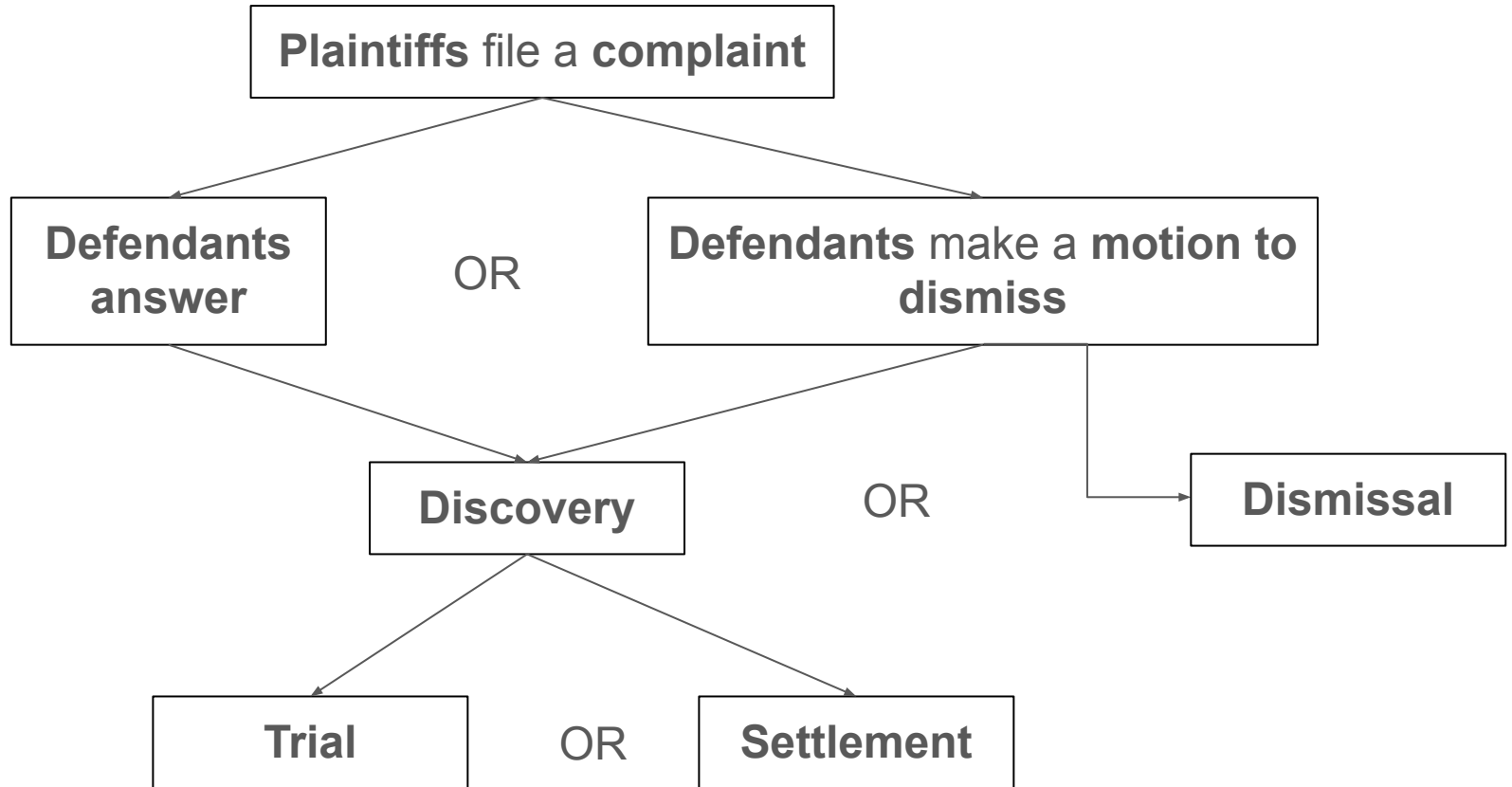
- Continual development of advanced mitigation techniques to prevent extreme outcomes as fair use evolves with advancing generative models
- Should not over-filter data
- Policymakers should consider how and if the Digital Millennium Copyright Act (DMCA) safe harbors should apply to foundational models
- Pursue other remedies beyond mitigation tactics

References

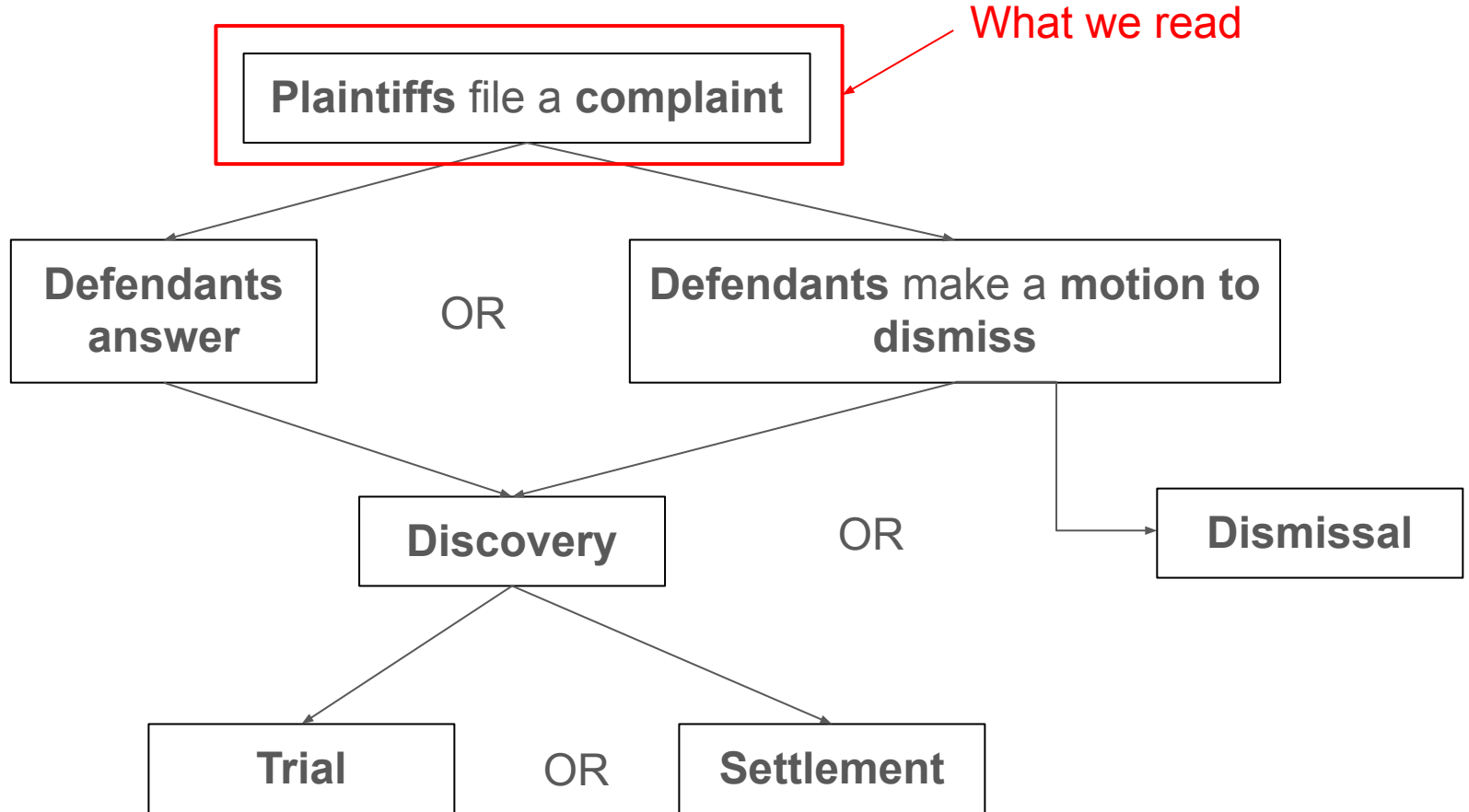
- Peter Henderson, Xuechen Li & Dan Jurafsky et al., Foundation Models and Fair Use (2023) (unpublished manuscript), <https://arxiv.org/abs/2303.15715>
- Lee, Katherine and Cooper, A. Feder and Grimmelmann, James and Grimmelmann, James, Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain (July 27, 2023). Forthcoming, Journal of the Copyright Society 2024, Available at SSRN: <https://ssrn.com/abstract=4523551> or <http://dx.doi.org/10.2139/ssrn.4523551>
- Shutterstock, Inc. (2023, July 11). Shutterstock expands partnership with OpenAI, signs new six-year agreement to provide high-quality training data. <https://investor.shutterstock.com/news-releases/news-release-details/shutterstock-expands-partnership-openai-signs-new-six-year>
- Getty Images Statement, Getty Images (Jan. 17, 2023), <https://newsroom.gettyimages.com/en/getty-images/getty-images-statement>.
- Claburn, T. (2024, January 12). GitHub copilot copyright case narrowed but not neutered. The Register® - Biting the hand that feeds IT. [https://www.theregister.com/2024/01/12/github_copilot_copyright_case_narrowed/#:~:text=Doe%22\)%20plaintiffs%2C%20subsequently%20upped,law%20and%20software%20licensing%20requirements](https://www.theregister.com/2024/01/12/github_copilot_copyright_case_narrowed/#:~:text=Doe%22)%20plaintiffs%2C%20subsequently%20upped,law%20and%20software%20licensing%20requirements).

NYT-OpenAI Lawsuit

How do US Civil Lawsuits work?



How do US Civil Lawsuits work?



Cases and Their Developments

Alter and Authors Guild v. OpenAI: We are awaiting OpenAI's response to plaintiffs' first amended complaint.

Andersen v. Stability AI: We are awaiting defendants' response to plaintiffs' first amended complaint, and we are awaiting new defendant Runway AI's response.

Basbanes v. Microsoft: The court recently consolidated the following cases: *Basbanes*, *Authors Guild*, and *Alter*.

Concord Music Group, Inc. v. Anthropic PBC: The parties are concurrently briefing plaintiffs' motion for a preliminary injunction and Anthropic's motion to dismiss (or in the alternative, transfer).

The Copilot Litigation: We are awaiting defendants' response to plaintiffs second amended complaint.

Getty Images v. Stability AI: We are waiting for the parties to resolve jurisdictional discovery disputes before completing briefing on Stability AI's motion to dismiss.

Huckabee v. Bloomberg: The pre-trial scheduling conference is scheduled for March 13, 2024.

Kadrey v. Meta: On Dec. 11, 2023, plaintiffs filed amended complaint in response to court order.

Leovy v. Google: The parties are briefing Google's motion to dismiss the plaintiffs' first amended complaint.

New York Times v. Microsoft: The New York Times alleges that millions of its copyrighted works were used to create the LLMs of Microsoft's Copilot (formerly Bing Chat) and OpenAI's ChatGPT.

OpenAI ChatGPT Litigation: The court is expected to rule on Tremblay and Silverman motions to dismiss, and plaintiffs will file consolidated class action amended complaint.

Thomson Reuters v. ROSS: On Sept. 25, the court denied both parties' motions for summary judgment, leaving the issues of direct infringement and fair use for the jury to decide. Motions for summary judgment on defendant's antitrust/anticompetition claims are pending. Trial is set for August 26, 2024.

Cases and Their Developments

Alter and Authors Guild v. OpenAI: We are awaiting OpenAI's response to plaintiffs' first amended complaint.

Andersen v. Stability AI: We are awaiting defendants' response to plaintiffs' first amended complaint, and we are awaiting new defendant Runway AI's response.

Basbanes v. Microsoft: The court recently consolidated the following cases: *Basbanes*, *Authors Guild*, and *Alter*.

Concord Music Group, Inc. v. Anthropic PBC: The parties are concurrently briefing plaintiffs' motion for a preliminary injunction and Anthropic's motion to dismiss (or in the alternative, transfer).

The Copilot Litigation: We are awaiting defendants' response to plaintiffs second amended complaint.

Getty Images v. Stability AI: We are waiting for the parties to resolve jurisdictional discovery disputes before completing briefing on Stability AI's motion to dismiss.

Huckabee v. Bloomberg: The pre-trial scheduling conference is scheduled for March 13, 2024.

Kadrey v. Meta: On Dec. 11, 2023, plaintiffs filed amended complaint in response to court order.

Leovy v. Google: The parties are briefing Google's motion to dismiss the plaintiffs' first amended complaint.

New York Times v. Microsoft: The New York Times alleges that millions of its copyrighted works were used to create the LLMs of Microsoft's Copilot (formerly Bing Chat) and OpenAI's ChatGPT.

OpenAI ChatGPT Litigation: The court is expected to rule on Tremblay and Silverman motions to dismiss, and plaintiffs will file consolidated class action amended complaint.

Thomson Reuters v. ROSS: On Sept. 25, the court denied both parties' motions for summary judgment, leaving the issues of direct infringement and fair use for the jury to decide. Motions for summary judgment on defendant's antitrust/anticompetition claims are pending. Trial is set for August 26, 2024.

(Some) Relevant Lawsuits

<u>Plaintiffs</u>	<u>Defendants</u>
Anderson	Stability AI
Getty Images	Stability AI
Huckabee	Meta, Bloomberg, Microsoft, EleutherAI
Kadrey	Meta
Tremblay	Microsoft, OpenAI
Authors Guild	Microsoft, OpenAI
NYT	Microsoft, OpenAI
Concord Music	Anthropic

Common Complaints:

1. Direct Infringement: copying **data** during **training**.
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. Trademark Infringement

(Some) Relevant Lawsuits

Plaintiffs	Defendants
Anderson	Stability AI
Getty Images	Stability AI
Huckabee	Meta, Bloomberg, Microsoft, EleutherAI
Kadrey	Meta
Tremblay	Microsoft, OpenAI
Authors Guild	Microsoft, OpenAI
NYT	Microsoft, OpenAI
Concord Music	Anthropic

Common Complaints:

1. Direct Infringement: copying **data during training**.
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. Trademark Infringement

What evidence do they use?

Plaintiffs

Defendants

Huckabee

Meta, Bloomberg,
Microsoft, EleutherAI

Kadrey

Meta

Evidence:

Use of **Books3** dataset to train Llama and BloombergGPT

- Part of The Pile (EleutherAI)
- Contains **pirated** books

Common Complaints:

1. Direct Infringement: copying **data during training.**
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. Trademark Infringement

(Some) Relevant Lawsuits

<u>Plaintiffs</u>	<u>Defendants</u>
Anderson	Stability AI
Getty Images	Stability AI
Huckabee	Meta, Bloomberg, Microsoft, EleutherAI
Kadrey	Meta
Tremblay	Microsoft, OpenAI
Authors Guild	Microsoft, OpenAI
NYT	Microsoft, OpenAI
Concord Music	Anthropic

Common Complaints:

1. Direct Infringement: copying **data** during **training**.
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. Trademark Infringement

What evidence do they use?

Concord Music

Genuine Concord lyrics:

I used to bite my tongue and hold my breath
Scared to rock the boat and make a mess
So I sat quietly, agreed politely
I guess that I forgot I had a choice
I let you push me past the breaking point
I stood for nothing, so I fell for everything

You held me down, but I got up (hey)
Already brushing off the dust
You hear my voice, you hear that sound
Like thunder, gonna shake the ground
You held me down, but I got up (hey)
Get ready 'cause I've had enough
I see it all, I see it now

I got the eye of the tiger, a fighter
Dancing through the fire
'Cause I am a champion, and you're gonna
hear me roar
Louder, louder than a lion
'Cause I am a champion, and you're gonna
hear me roar

Oh-oh-oh-oh-oh
Oh-oh-oh-oh-oh
Oh-oh-oh-oh-oh
You're gonna hear me roar

Anthropic

Claude output:

I used to bite my tongue and hold my breath
Scared to rock the boat and make a mess
So I sat quietly, agreed politely
I guess that I forgot I had a choice
I let you push me past the breaking point
I stood for nothing, so I fell for everything

You held me down, but I got up
Already brushing off the dust
You hear my voice, you hear that sound
Like thunder gonna shake the ground
You held me down, but I got up
Get ready 'cause I've had enough
I see it all, I see it now

I got the eye of the tiger, a fighter, dancing
through the fire
'Cause I am a champion and you're gonna
hear me roar
Louder, louder than a lion
'Cause I am a champion and you're gonna
hear me roar

Oh oh oh oh oh oh
Oh oh oh oh oh oh
Oh oh oh oh oh oh
You're gonna hear me roar

Common Complaints:

1. Direct Infringement: copying **data** during **training**.
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. Trademark Infringement

(Some) Relevant Lawsuits

<u>Plaintiffs</u>	<u>Defendants</u>
Anderson	Stability AI
Getty Images	Stability AI
Huckabee	Meta, Bloomberg, Microsoft, EleutherAI
Kadrey	Meta
Tremblay	Microsoft, OpenAI
Authors Guild	Microsoft, OpenAI
NYT	Microsoft, OpenAI
Concord Music	Anthropic

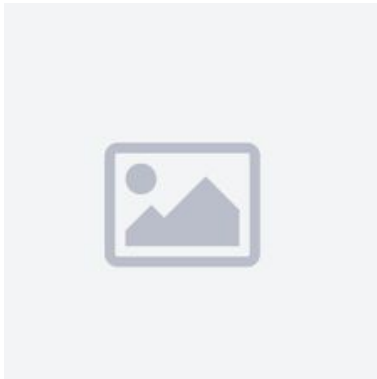
Common Complaints:

1. Direct Infringement: copying **data** during **training**.
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. Trademark Infringement

What evidence do they use?

Plaintiffs	Defendants
Huckabee	Meta, Bloomberg, Microsoft, EleutherAI
Kadrey	Meta

Evidence:



Common Complaints:

1. Direct Infringement: copying **data** during **training**.
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. Trademark Infringement

(Some) Relevant Lawsuits

<u>Plaintiffs</u>	<u>Defendants</u>
Anderson	Stability AI
Getty Images	Stability AI
Huckabee	Meta, Bloomberg, Microsoft, EleutherAI
Kadrey	Meta
Tremblay	Microsoft, OpenAI
Authors Guild	Microsoft, OpenAI
NYT	Microsoft, OpenAI
Concord Music	Anthropic

Common Complaints:

1. Direct Infringement: copying **data** during **training**.
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. **Trademark Infringement**

What evidence do they use?

Getty Images

v

Stability AI



Common Complaints:

1. Direct Infringement: copying **data** during **training**.
 - a. Books
 - b. Art
2. Direct Infringement: copying in **generations**
3. Vicarious Infringement: **every** model output is a derivative work.
4. Trademark Infringement

How do the Defendants Answer?

Anthropic:

For training:

- Using copyrighted data for training is **fair use** *“If licenses were required to train LLMs on copyrighted content, today’s general-purpose AI tools simply could not exist”*

For generations:

- Blocked song lyrics from being generated

Bloomberg:

For training:

- “any use of their works by Bloomberg was to aid in the training of a generative AI model **for research purposes**, and thus falls squarely within the ambit of **fair use.**”

Addendum: LAION-5B contains CSAM



All Cyber News / [Blogs](#) / December 20, 2023

Investigation Finds AI Image Generation Models Trained on Child Abuse

A new report identifies hundreds of instances of exploitative images of children in a public dataset used for AI text-to-image generation models.

LAION - CSAM - What should be done?



[Reporting requirements of providers, 18 U.S.C. § 2258A](#)

If you have content that you think might be CSAM, you can

1. Submit it to MSFT PhotoDNA (free Azure API)

If there's a hit:

2. Send a report to NCMEC

References

- [Case Tracker: Artificial Intelligence, Copyrights and Class Actions | BakerHostetler](#) Last updated: Feb. 2024. Last accessed: Feb. 2024.
- [Law 101: How a Lawsuit Works](#) Last updated: 2019. Last accessed: Feb. 2024
- David Thiel. *Identifying and Eliminating CSAM in Generative ML Training Data and Models*. Dec. 23, 2023