NO PREP

Stolen Voices and Stolen Art

Copyright in the AI Age

Includes: Student reading, case study, differentiated questions, teacher key

MORE LESSONS LESS PLANNING

MLLP

FREE TODAY · FREE TOMORROW · FREE FOREVER

Name: Period: Date:

Stolen Voices and Stolen Art: Copyright in the AI Age

The Training Data Question

Every AI system that generates text, images, music, or code learned to do so by analyzing existing human-created work. Large language models studied billions of pages of writing. Image generation models were trained on hundreds of millions of photographs, illustrations, and artworks. Music generation AI learned from recordings. Code generation AI was trained on millions of open-source and publicly available programming files. The AI did not create in a vacuum — it learned by absorbing what human beings had already made, and it now uses those learned patterns to generate new outputs. The central legal and ethical

Name: **Period:** **Date:**

Stolen Voices and Stolen Art: Copyright in the AI Age

The Training Data Question

Every AI system that generates text, images, music, or code learned to do so by analyzing existing human-created work. Large language models studied billions of pages of writing. Image generation models were trained on hundreds of millions of photographs, illustrations, and artworks. Music generation AI learned from recordings. Code generation AI was trained on millions of open-source and publicly available programming files. The AI did not create in a vacuum: it learned by absorbing what human beings had already made, and it now uses those learned patterns to generate new outputs. The central legal and ethical question this raises is: did the companies that trained these AI systems have the right to use all of that material? For most of the works in AI training datasets, the answer was assumed by default to be yes, because the material was publicly available on the internet. But copyright law does not make a simple exception for publicly accessible content. A photograph posted on a personal website, an artist's illustration published on a portfolio page, or a musician's recording shared on a streaming platform is still protected by copyright. The person who created it retains rights over how it is used. Those rights have never explicitly been waived for AI training purposes.

Artists Fight Back

In January 2023, a group of visual artists (including Sarah Andersen, Kelly McKernan, and Karla Ortiz) filed a lawsuit against Stability AI, Midjourney, and DeviantArt, alleging that these AI image generation companies had used their artwork without permission to train their systems. The lawsuit argued that the AI tools could even produce images in the distinctive styles of individual artists, and that style, while technically not copyrightable, could be so closely replicated as to effectively compete with and undercut the original artist's market. In the same

month, Getty Images filed a separate lawsuit against Stability AI, alleging that the company had scraped millions of Getty's licensed photographs to train its image generator. Getty's lawsuit included a striking exhibit: the AI's outputs sometimes showed visible, degraded versions of Getty's watermark, evidence that specific, copyrighted images had been directly incorporated into training. Stability AI disputed the claims and argued that AI training constitutes "fair use" under copyright law. The litigation is ongoing as of 2026, and the outcomes of these cases will significantly shape what AI companies are allowed to do with existing copyrighted material.

The Voice Cloning Problem

Music and voice present a specific and particularly visceral dimension of the AI copyright problem. AI voice cloning tools can analyze a recording of a person speaking or singing and produce new audio that sounds convincingly like that person saying or singing anything the user inputs. These tools have been used to produce fake songs attributed to real artists: a song in the voice of Drake and The Weeknd generated by AI was briefly released on streaming platforms in 2023 before being removed, generating millions of streams and raising urgent questions about whether such content was legal. Voice cloning also poses serious non-copyright concerns: it can be used to produce nonconsensual intimate audio content, political disinformation, fraud, and harassment. But the copyright dimension is significant in its own right. A performer's voice is part of their artistic identity and economic livelihood. When an AI system can produce convincing imitations of that voice at scale, it threatens not just artistic integrity but the economic foundation of a performer's career. In 2023, major record labels (including Universal Music Group, Sony Music, and Warner Music Group) sent letters to AI companies and streaming platforms demanding that AI-generated content mimicking their artists be removed.

Code, Books, and the "Opt-Out" Debate

The copyright controversy extends well beyond images and music. In 2023, a group of authors including George R.R. Martin and John Grisham, through the Authors Guild, filed a class action lawsuit against OpenAI, alleging that the company had used their books — without permission or compensation — to train ChatGPT. OpenAI has maintained that training on publicly available text constitutes fair use. The New York Times filed a separate lawsuit against OpenAI and Microsoft in December

2023, alleging that millions of its articles were used in training and that the AI could reproduce those articles in ways that directly compete with the Times as a news source. One proposed solution is an "opt-out" system: creators who do not want their work used for AI training can register to have their work excluded from future datasets. Critics argue this gets the consent framework exactly backwards — the appropriate standard should be opt-in, with creators affirmatively choosing to allow their work to be used. The burden should be on the companies seeking to profit from that work, not on the individual creators who produced it. The debate about training data consent is at the heart of the ongoing legal and regulatory battles, and no clear legal resolution exists yet in the United States as of 2026.

Compensation Models and the Future

Some AI companies and creative industry groups have begun exploring compensation models. Adobe, which trains some of its AI tools on its own licensed stock image library, has argued that this represents a more ethical approach and has offered payments to stock contributors whose work is used in training. Shutterstock reached a deal with OpenAI in 2023 to license its image library for training and established a fund to compensate contributors. The music industry has begun negotiating licensing agreements with some AI music generation companies. These arrangements suggest that compensation models are possible — but critics note that most existing AI training datasets were assembled before these conversations began, and companies that have already trained on billions of uncompensated works are not being required to pay retroactively. The fundamental question — whether AI training on copyrighted material constitutes fair use — will ultimately be decided by courts. Copyright law in the United States was written long before AI training existed as a practice, and the courts are being asked to apply its principles to a situation its authors never anticipated. The answer will have enormous economic consequences for AI companies, the creative industries, and the millions of artists, writers, musicians, and programmers whose work is already inside these models, whether they agreed to be there or not. The legal battles over AI and copyright are ultimately battles over who owns the past. If the work of millions of human creators — writers, artists, musicians, programmers — can be absorbed into commercial AI systems without consent or compensation, then the value those creators generated accrues entirely to the companies that trained on it. The creative labor of a lifetime

becomes, in effect, a free input into someone else's product. How courts resolve this question will determine not only the economics of the creative industries but what incentives remain for human creativity in a world where AI can produce plausible imitations of almost any style on demand.

Key Information: AI and Copyright

Detail	Information
Core legal question	Whether using copyrighted material to train AI constitutes fair use
January 2023 artist lawsuit	Artists sued Stability AI, Midjourney, and DeviantArt for training on their work without consent
Getty Images lawsuit	Alleged Stability AI scraped millions of licensed photos; AI outputs showed degraded watermarks
Drake/Weeknd AI song (2023)	Fake AI-generated song briefly on streaming platforms; removed after record label pressure
Authors Guild lawsuit (2023)	Sued OpenAI for using books by George R.R. Martin, John Grisham, and others without permission
NYT lawsuit (Dec. 2023)	Sued OpenAI and Microsoft over use of millions of articles in AI training
Opt-out vs. opt-in debate	Companies propose opt-out; critics say consent should be opt-in by default
Compensation examples	Adobe, Shutterstock reached licensing/payment deals; but past training data uncompensated

Case Study: Lena's Illustration Portfolio

Lena has been a freelance illustrator for six years. Her work — detailed botanical illustrations with a distinctive color palette and linework — has appeared in cookbooks, magazines, and on product packaging. She uploads her portfolio to her website and to design community platforms so that potential clients can find her. Last year, a client contacted her asking for botanical illustrations. Then mentioned they had been using an AI tool and wondered if she could match its style. They sent her an AI-generated sample. She recognized the style immediately. It looked like her work. She asked the AI company for information about their training data. They did not respond. She found a lawyer who told her that proving her specific images were used in training would be extremely difficult and expensive. "I built this style over six years," she tells a colleague. "It's how people know my work. And now a machine can do it in seconds and undercut my prices. And I never said yes."

1. The reading identifies two distinct types of harm that can come from AI training on artists' work. Using Lena's situation, describe both types of harm.

2. Lena's lawyer says proving her specific images were used would be extremely difficult. Based on the reading, what piece of evidence in the Getty case showed that specific images had been directly used in training?

Multiple Choice

1. What argument do AI companies make to justify training on copyrighted material without explicit permission?

a) Copyright protections expire after ten years for any work published digitally online

b) Training on publicly available material constitutes "fair use" under copyright law

c) Copyright law does not apply to material hosted on foreign-owned servers

d) Artists implicitly consent to AI training by publishing their work online

2. What did the AI-generated song involving Drake and The Weeknd demonstrate in 2023?

a) That record labels support the use of AI to produce new music by their artists

b) That AI-generated content mimicking real artists can reach massive audiences before being removed

c) That AI music generation tools cannot yet replicate distinctive vocal styles

d) That streaming platforms have effective systems for detecting and quickly blocking any AI content

3. According to the reading, what is the core criticism of "opt-out" consent frameworks for AI training?

a) Companies would need to obtain prior government approval before adding a creator to an opt-out list

b) Opt-out lists are too expensive for small creators to register with

c) Opt-out frameworks have already been banned in the European Union and Canada

d) The burden should be on companies to get affirmative consent, not on creators to ask for exclusion

Matching

Write the letter of the correct definition on the line next to each term.

Term	Definition
1. ___ Fair use	A) AI technology that replicates a person's voice from recordings to generate new audio
2. ___ Voice cloning	B) A legal doctrine allowing limited use of copyrighted material without permission under certain conditions
3. ___ Opt-in consent	C) Organization that filed a class action lawsuit against OpenAI on behalf of authors in 2023
4. ___ Training data	D) A consent model in which creators must actively agree before their work can be used
5. ___ Authors Guild	E) The existing human-created content used to teach an AI system to produce outputs

Fill-in-the-Blank

1. Getty Images' lawsuit against Stability AI included evidence that AI outputs sometimes showed degraded versions of Getty's _____ .

2. The New York Times filed its lawsuit against OpenAI and Microsoft in _____ 2023.

3. Adobe argues that training AI tools on its own _____ image library represents a more ethical approach.

Short Answer

Answer each question in 2-4 complete sentences.

1. Why does the opt-in versus opt-out debate matter, and which side has the stronger argument in your view? Use evidence from the reading to support your position.

2. Explain why voice cloning raises concerns beyond copyright law, according to the reading.

3. Why has it been difficult to resolve AI copyright questions through existing copyright law?