

Generative AI & Artistic Rights



ICAAD
Human Rights Innovation



International Center for Advocates Against Discrimination (ICAAD) is a human rights advocacy center working at the intersection of legal innovation and design justice to create evidence-based programs with governments and local communities to combat structural discrimination. By leveraging multidisciplinary teams and taking an integrated approach, we are able to improve resilience, safety, and equity across systems.

For any inquiries, further information, or to discuss any aspect of this report, please feel free to reach out to us at info@icaad.ngo. Your feedback and questions are highly valued.

AUTHOR

Navreet Kaur, ICAAD Intern

Summary

Evolving generative AI technologies have begun to transform the art industry. Some of these innovations¹ have proven useful in aiding artists with their craft, but the advancements bring a legal gray area, raising concerns for many artists. There are questions regarding the legality of the data that companies use to train these expansive AI models, and many artists have already experienced the negative impacts of AI. Although there have been proposed solutions, this area is constantly evolving and lacks legal precedent, creating uncertainty for artists and developers alike. This paper explores the legal, practical, and technological responses to these issues, emphasizing the need for transparent data practices and robust licensing agreements to protect artists' rights and creativity.

Development of Generative AI and Art

In 2020, a groundbreaking paper² written by theoretical physicist, Jared Kaplan, highlighted that the more data used to train an AI model, the better it would perform. Prior to the publication of this paper, AI models were trained using relatively little training data. However, Kaplan's paper made it clear that to perform efficiently and accurately, these models must be trained using extremely large datasets, often consisting of copyrighted materials scraped³ from the internet. Larger datasets resulted in AI models that could create detailed outputs faster than most humans could think.

These technological advancements have revolutionized the industry at a cost to artists. Companies prioritize rapid development and outperforming competitors over ethical concerns and the rights of artists because it would "take too long" to negotiate⁴ licenses. One lawyer at Meta even warned⁵ of ethical concerns surrounding the taking of intellectual property from artists but was ignored. In 2023, Google revised its privacy policy,⁶ enabling them to "use publicly available information to help train Google's AI models." These new terms were strategically released⁷ on the Fourth of July weekend with hopes that people would be too busy with the holiday to notice.

Common Crawl

Currently, AI models are trained using extremely large data sets. LAION, Large-scale Artificial Intelligence Open Network, a German non-profit that creates open-sourced artificial intelligence models and datasets, uses image data from Common Crawl,⁸ a public web archive that gathers image and text information from over 3 billion websites. Although most web archive pages⁹ do not make their content accessible for bulk data mining due to copyright concerns, Common

¹ Lucy Handley, "Part scary, part exciting: How artists are using AI in their work," CNBC (Apr. 1, 2024), <https://www.cnbc.com/2024/04/01/generative-ai-in-art-how-artists-are-using-it-or-not.html#:~:text=Generative%20AI%20can%20create%20images,the%20potential%20of%20generative%20AI>.

² Jared Kaplan et al., "Scaling Laws for Neural Language Models," at 3 (Jan. 23, 2020), <https://arxiv.org/pdf/2001.08361>.

³ Wen Wang et al., "Generative AI and Artists: Consumer Preferences for Style and Fair Compensation," at 1 (Sept. 1, 2023), <https://drive.google.com/file/d/1UpbvL5PaxTWHQNvtGCvBDpbRB2WYLFuU/view>.

⁴ Cade Metz et al., "How Tech Giants Cut Corners to Harvest Data for A.I.," NY Times (Updated Apr. 6 2024), <https://www.nytimes.com/2024/04/06/technology/tech-giants-harvest-data-artificial-intelligence.html>.

⁵ *Id.*

⁶ *Id.*

⁷ *Id.*

⁸ <https://commoncrawl.org/get-started>.

⁹ Kalev Leetaru, "Common Crawl And Unlocking Web Archives For Research," Forbes (Sept. 28, 2017), <https://www.forbes.com/sites/kalevleetaru/2017/09/28/common-crawl-and-unlocking-web-archives-for-research/>.

Crawl allows users to freely download data. Their goal¹⁰ is to make the volume and quality of data that is available to large corporations accessible to all.

Common crawl does not store images, as it collects “raw web page data, metadata extracts, and text extracts.”¹¹ This data is then stored in web archive transformation files, which use WebAssembly text format to store data about images. These files are then stored and published by CommonCrawl in a format that allows LAION to evaluate the metadata of websites without copying or downloading the images themselves. After processing, LAION released a five-billion image-text pair dataset called LAION-5B in 2022, “the largest freely available image-text dataset.”¹² This dataset was used by companies such as Stability AI to train their models.

This unique process makes it difficult to analogize with prior legal cases¹³ where the images themselves were being copied or stored. Furthermore, models such as Dall-E 2 and Stable Diffusion do not search their training sets for specific images to match a text prompt like Google does. Instead, the programs learn what images need to look like to match textual words and phrases to create an image that satisfies the prompts they are given. This machine-driven process makes it difficult to prove direct copying because the generative AI does not find and copy or collage¹⁴ images using image files, though it does use the alt-attributes description of the image file. Many artists have even found their names used for AI prompts, resulting in thousands of works nearly identical to the originals. This raises concerns, especially considering that AI is capable of generating artwork far more quickly than humans. Many defenders of AI have dismissed these concerns, stating that it is very difficult to replicate a pre-existing image using generative AI, and that many of the images used do not have copyrightable elements.

There are limitations to Common Crawl’s data collection because many large domains, such as Facebook and the New York Times, have prohibited¹⁵ Common Crawl from using their information. In 2023, The New York Times reached out to the Common Crawl Foundation to get its content removed¹⁶ from the dataset. Common Crawl complied with their request and agreed not to scrape any additional copyrighted content from The New York Times in the future. While these limitations prevent Common Crawl from obtaining a representative sample of the web, they also protect against copyright infringement claims and related lawsuits. Additionally, Common Crawl only collects the HTML code of the pages used rather than full copies of domains. Common Crawl also suggests web domain administrators add code to their websites robots.txt file¹⁷, to instruct their web crawler not to crawl their website.

¹⁰ <https://commoncrawl.org/mission>.

¹¹ <https://commoncrawl.org/overview>.

¹² Anthony Alford, *LAION Releases Five Billion Image-Text Pair Dataset LAION-5B*, May 17, 2022, <https://www.infoq.com/news/2022/05/laion-5b-image-text-dataset/>

¹³ Michael D. Murray, “Generative AI Art: Copyright Infringement and Fair Use,” at 272, <https://drive.google.com/file/d/1G5B5BHkkieiXEBwsXGcNcq4cXrG88TSi/view>.

¹⁴ *Id.* at 23.

¹⁵ Stefan Baack, “Common Crawl’s Impact on Generative AI,” Mozilla (Feb. 6, 2024), <https://foundation.mozilla.org/en/research/library/generative-ai-training-data/common-crawl/>.

¹⁶ Alistair Barr & Kali Hays, “The New York Times got its content removed from one of the biggest AI training datasets. Here's how it did it.,” Business Insider (Nov. 8, 2023), <https://www.businessinsider.com/new-york-times-content-removed-common-crawl-ai-training-dataset-2023-11>.

¹⁷ <https://commoncrawl.org/ccbot>.

Synthetic Data

The New York Times¹⁸ reports that tech companies could run out of high-quality data used to train their models as soon as 2026 because they are using the data faster than it is being produced. Companies such as Open AI are preparing for the looming shortage of data by developing¹⁹ synthetic data for the models to be trained on. Synthetic data enables²⁰ AI systems to learn from their supposed own creations. This allows companies to rapidly generate large datasets without the restrictions that come with organic data because individual privacy is protected and it prevents companies from exposing personal or copyrighted material. However, synthetic data is not as reliable²¹ as organic data and often does not emulate real-world conditions. Additionally, synthetic data may exaggerate²² imperfections in the original datasets they are trained on and perpetuate or amplify its own mistakes, resulting in a lower-quality data set over time. This creates a risk of AI model degradation if it is not periodically synched²³ to real-world data. To combat this, Open AI and other companies are exploring how two different AI models may work together, where one system produces data and the second filters out the mistakes. While synthetic data will prove to be useful in the future, it does not address present-day concerns.

Preventing the Use of Data

Concerns over copyright infringement have led to the development of websites such as “Have I Been Trained,”²⁴ where artists can search for their work in popular AI training datasets. The website allows users to find exact matches using image captions, artist names, or even a description of the image. A quick search on the site reveals that it uses everything from Van Gogh to elementary school artwork. The website also has a “Do Not Train Registry.”²⁵ allowing users to claim their domain and set permissions on the usage of their images, with the goal of giving individuals more control over their creative work. This registry does not remove data from already trained models, nor is it binding upon AI companies, as they must agree to honor the do not train registry. The list²⁶ of partner organizations that have agreed is growing and now includes HuggingFace (the largest repository of models and datasets) and Stability AI.

The Glaze Project²⁷ is a research effort that aims to develop technical tools to protect human creatives against infringing uses of generative AI. Its main goal is to disrupt unauthorized AI training on artists' works and allow them to retain agency and control over the use of their work products.

Style mimicry enables AI models to fine-tune images to the style of a specific artist. This can lead to the loss of commissions and income, as well as the dilution of their style, brand, and reputation that took years to develop. This has led to the demoralization of many young and aspiring artists, causing plummeting student enrollment. At Indiana University Bloomington, enrollment in the Arts

¹⁸ “How Tech Giants Cut Corners to Harvest Data for A.I.,” NY Times, *supra* 4.

¹⁹ *Id.*

²⁰ *Id.*

²¹ Ivan Kot, “The Pros and Cons of Synthetic Data,” Dataversity (Dec. 27, 2021), <https://www.dataversity.net/the-pros-and-cons-of-synthetic-data/>.

²² Kim Bozzella, “The Pros And Cons Of Using Synthetic Data For Training AI,” (Nov. 20, 2023), <https://www.forbes.com/sites/forbestechcouncil/2023/11/20/the-pros-and-cons-of-using-synthetic-data-for-training-ai/>.

²³ *Id.*

²⁴ <https://haveibeen trained.com/>.

²⁵ HaveIBeenTrained, “Frequently Asked Questions,” <https://haveibeen trained.com/faq>.

²⁶ <https://spawning.ai/have-i-been-trained>.

²⁷ <https://glaze.cs.uchicago.edu/>.

and Sciences plummeted²⁸ from 9066 students in 2014 to 7008 in 2022. Low enrollment has caused many art schools to shut down,²⁹ including The University of the Arts,³⁰ which was established in 1876.

Glaze works to combat this by understanding³¹ the models that are training on human art and using machine learning algorithms to compute minimal changes to the artwork. These changes are nearly invisible to human eyes but create a significant difference in the output of an AI, protecting artists from style mimicry. It is nearly impossible to override Glaze because, unlike a watermark or steganography, it is a new dimension embedded within the artwork that a human cannot see, but an AI model can see, and is impossible to interrupt without an attacker knowing the specific dimensions. The Glaze project also released Nightshade,³² a tool that helps protect artists against their content being used for model training without their permission. Unlike opt-out lists and do not scrape directives, which can be difficult to enforce, Nightshade is an offensive tool that distorts feature representations inside generative AI image models to make the images useless for model training. The Glaze project is currently working on an integrated release of both Nightshade and Glaze to make the tools more accessible for artists.

Opt-In Data Collection

Companies should be required to notify³³ individuals before their data is used for AI training. The current trend of requiring data owners to opt out of data collection is not effective, as it requires knowledge of the data that is going to be used. Furthermore, companies often make it difficult to opt out of data collection, “and even where it is possible, many people don’t have a “clear idea” about the permissions they’ve agreed to or how data is being used.”³⁴ As a result, artists are often not aware that their data has been used until it is too late, and it is nearly impossible³⁵ to make an AI model “forget” what it has learned. A more equitable option is to opt into data collection for AI training, as this ensures the data owner is aware of the data being used.

Meta recently announced that they would be using public Facebook and Instagram posts to train³⁶ its generative AI models. For United States users, there is no way to stop³⁷ Meta AI from

²⁸ Aubrey Wright, “As arts and humanities enrollment declining, IU advocates say degrees still valuable,” Indiana Public Media (May. 10, 2024) <https://indianapublicmedia.org/news/as-arts-and-humanities-enrollment-declining-iu-advocates-say-degrees-still-valuable.php>.

²⁹ The Associated Press, “University of the Arts abruptly announces June 7 closure, vows to help students transfer,” The Morning Call (June. 1, 2024) <https://www.mcall.com/2024/06/01/university-of-the-arts-abruptly-announces-june-7-closure-vows-to-help-students-transfer/>.

³⁰ University of the Arts, [About UArts](https://www.uarts.edu/about#:~:text=Established%20in%201876%2C%20the%20university,College%20of%20the%20Performing%20Arts), <https://www.uarts.edu/about#:~:text=Established%20in%201876%2C%20the%20university,College%20of%20the%20Performing%20Arts>.

³¹ Glaze, “What is Glaze?,” <https://glaze.cs.uchicago.edu/what-is-glaze.html>.

³² Nightshade, “What is Nightshade?,” <https://nightshade.cs.uchicago.edu/whatis.html>.

³³ Kathryn Shih, “Protecting Proprietary Data Rights in the AI Era,” Forgepoint (Apr. 2, 2024) <https://forgepointcap.com/perspectives/protecting-proprietary-data-rights-in-the-ai-era/>.

³⁴ Matt Burgess & Reese Rogers, “How to Stop Your Data From Being Used to Train AI,” Wired (Apr. 10, 2024) <https://www.wired.com/story/how-to-stop-your-data-from-being-used-to-train-ai/>.

³⁵ *Id.*

³⁶ Melissa Heikkilä, “How to opt out of Meta’s AI training,” MIT Technology Review (Jun. 14, 2024) <https://www.technologyreview.com/2024/06/14/1093789/how-to-opt-out-of-meta-ai-training/>.

³⁷ *Id.*

learning from public social media posts. Additionally, Meta noted in a statement³⁸, “While we don’t currently have an opt-out feature, we’ve built in-platform tools that allow people to delete their personal information from chats with Meta A.I. across our apps.”³⁹ Those in the European Union, United Kingdom, and Switzerland were given the option⁴⁰ to opt-out due to their strict data privacy laws. In 2018, Europe replaced old data protection rules with updated rules⁴¹ that reflect the technological advancements that have been emerging. General Data Protection Regulation (“GDPR”) places limits on what organizations can do with personal data with the potential for large fines and reputational damage⁴² for those who breach the rules. GDPR⁴³ protects personal data that allows a living person to be directly or indirectly identified, including pseudonymized data.

Copyright Protections Before AI

Copyright laws protect original works that are independently created and sufficiently creative. “Creativity can be demonstrated in a variety of ways and reflects artistic choices like the subject matter, composition, depiction, and the use of the elements of design.”⁴⁴ The Visual Artists Rights Act (“VARA”), a U.S. law, prevents others from “intentionally distorting, mutilating, or modifying artwork in a way that dishonors the artist’s reputation.”⁴⁵ Although copyright protection exists from the moment the work is fixed,⁴⁶ registering with the U.S. Copyright Office offers additional protections.

A derivative work is derived from one or more already existing works. Translation is considered derivative work, and to be copyrightable,⁴⁷ it must incorporate some or all of the preexisting material. “A “compilation”⁴⁸ is a work formed by the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship.” Compilations of data and preexisting works are also copyrightable⁴⁹ if the materials are selected or arranged in a way that the resulting work is considered a new work. Only the copyright owner of a work has the right to prepare or authorize⁵⁰ another to create an adaptation of their work.

³⁸ Jesus Jiménez, “Can I Opt Out of Meta’s A.I. Scraping on Instagram and Facebook? Sort Of.,” NY Times (Jun. 7, 2024) <https://www.nytimes.com/2024/06/07/technology/meta-ai-scraping-policy.html>.

³⁹ *Id.*

⁴⁰ *Id.*

⁴¹ Matt Burgess, “What is GDPR? The summary guide to GDPR compliance in the UK,” Wired (Mar. 24, 2020) <https://www.wired.com/story/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018/>.

⁴² *Id.*

⁴³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1552662547490&uri=CELEX%3A32016R0679>.

⁴⁴ US Copyright Office, “What Visual and Graphic Artists Should Know About Copyright,” <https://www.copyright.gov/engage/visual-artists/>.

⁴⁵ Minnesota Lawyers for the Arts, “Visual Artwork and Copyright Protection,” <https://springboardforthearts.org/wp-content/uploads/2014/06/Visual-Art-and-Copyright.pdf>.

⁴⁶ In the context of copyright law, the term “fixed” refers to the work being captured in a tangible form, such that it can be perceived, reproduced, or otherwise communicated for more than a short period of time.

⁴⁷ US Copyright Office, “Copyright in Derivative Works and Compilations,” Library of Congress (Jul. 2020), <https://www.copyright.gov/circs/circ14.pdf>.

⁴⁸ <https://www.copyright.gov/title17/92chap1.html>.

⁴⁹ *Id.*

⁵⁰ US Copyright Office, “Copyright in Derivative Works and Compilations.”

Legal Action

Because it is difficult to find direct evidence regarding actual copying, the law looks at the “likelihood of copying”⁵¹ when analyzing claims. This test examines whether the defendant had access to the work and if there is substantial similarity between the works. For a generative AI copyright claim, it is possible to satisfy this requirement if the specific work is included in a dataset used to train a generative AI model. However, potential access is not proof of actual copying, so the next step of the analysis looks to “substantial similarity,”⁵² which requires the court to determine if the works are extrinsically similar because they contain substantially similar ideas subject to copyright protection and whether the works are intrinsically similar because they express the ideas in a substantially similar manner from the perspective of the intended audience. This means that to establish copyright infringement, the defendant must have generated a work that can be compared side by side with the allegedly infringed work, a process that is far removed from the actualities of AI training.

Due to the lack of precedent, courts refer to past cases⁵³ regarding transformative fair use in the context of non-expressive copying and copy-reliant technologies. Typically, this refers to incidental or temporary processing of data from copyrighted works out of necessity. Generative AI technology companies argue that AI foundation models used to train generative AI systems are a form of copy-reliant, non-expressive copying. In the past, courts have analyzed the use of search engines to crawl the web to scrape images and make exact copies of copyrighted images as seen in *Kelly v. Arriba Soft Corp.* and *Perfect Inc. v. Amazon.com, Inc.* In both cases, image owners sued the companies for copying images for thumbnails, storing them on servers, and displaying the images on their search engines. The courts found both instances to be fair uses because the images were merely used as visual cues to sources of information, not as aesthetic objects. One major factor of the fair use analysis in *Kelly* was that the images were not used for an artistic purpose and did not replace the need for the originals.

NO FAKES Act

The Nurture Originals, Foster Art, and Keep Entertainment Safe (“NO FAKES”) Act⁵⁴ of 2023 is a bipartisan proposal aiming to protect⁵⁵ individuals from unauthorized recreations of faces, names, and voices by generative AI. The draft legislation has three components:⁵⁶ “Hold individuals and companies accountable if they produce an unauthorized digital replica of an individual in a performance; Hold platforms liable for hosting an unauthorized digital replica if the platform has knowledge of the fact that the replica was not authorized by the individual depicted; and Exclude certain digital replicas from coverage based on recognized First Amendment protections.”⁵⁷ These consequences create a strong disincentive for unauthorized use. Whereas prior common law only protected identities with commercial value, the NO FAKES Act does not require any commercial value in the material used, punishing those who use data with consent. Furthermore,

⁵¹ Michael D. Murray, “Generative AI Art: Copyright Infringement and Fair Use,” at 269.

⁵² *Id.*

⁵³ *Id.* at 272.

⁵⁴ https://www.coons.senate.gov/imo/media/doc/no_fakes_act_draft_text.pdf.

⁵⁵ Senators Chris Coons, Marsha Blackburn, Amy Klobuchar, Thom Tillis, “Nurture Originals, Foster Art, and Keep Entertainment Safe,” https://www.coons.senate.gov/imo/media/doc/no_fakes_act_one_pager.pdf.

⁵⁶ *Id.*

⁵⁷ *Id.*

damages⁵⁸ are not limited to commercial gains or losses, and the proposed bill allows for \$5,000 in damages for each violation and punitive damages.

Although the proposed legislation⁵⁹ does not directly address visual artwork that does not represent a person's image or likeness, it opens the door for additional protections for visual artists because many of the concerns are similar. Tahliah DebreTT Barnett, a singer and songwriter, expressed her concerns at the Senate Judiciary hearing,⁶⁰ stating, “My art is a canvas on which I paint my identity and the sustaining foundation of my livelihood. It is the very essence of my being. Yet this is under threat. AI cannot replicate the depth of my life journey, yet those who control it hold the power to mimic the likeness of my art, replicate it, and falsely claim my identity and intellectual property. This prospect threatens to rewrite and unravel the fabric of my very existence.”

Licensing

The volume of data needed to train AI models is so large⁶¹ that “even collective licensing can't work.” Another possible approach to combat the ethical concerns raised is compensating artists based on the value added. However, this approach is subject to a set of preconditions, such as the common usage of an artist's name and the willingness of users to pay. A study⁶² found that many AI users are willing to pay for the products they use if a royalty is given to contributing artists. Additionally, users are the most willing to pay when contributing artists are compensated per use instead of a one-time payment. This is explained by a consumer desire for ethically sourced products, allowing marketers to pass along increased costs to consumers while compensating artists. Many large content licensors, such as LexisNexis and Bloomberg, offer⁶³ licensed commercial bulk feeds that are designed to support data mining access. This ensures that content owners are paid royalty fees for the material that is used.

Dataset Shop⁶⁴ offers licensing options for individuals seeking data for generative AI training for visual synthetic media. There are options for a one-time license, which includes access to existing datasets, or for an annual subscription, which secures access to current datasets and updates. The images in the datasets are “either fully owned by vAIsual Inc or have been acquired under a legal license from third party providers.”

A licensing agreement should contain several use restrictions to protect artists. First, it should clearly define the permitted uses of data, such as only using the training data to train an AI model. There should also be a section emphasizing the confidentiality of the training data to prevent the data from being disclosed to third parties. Licenses from DefinedAi and Microsoft are a couple of examples that contain these provisions.

⁵⁸ Jennifer A. Kenedy & Jordan Rutledge, “Locke Lord QuickStudy: The NO FAKES Act: With Proposed Bill, Congress Set to Protect Against Unauthorized Digital Replicas of Faces, Names and Voices,” Locke Lord (Oct. 16, 2023) https://www.lockelord.com/newsandevents/publications/2023/10/no-fakes-act#_ftn5.

⁵⁹ https://www.coons.senate.gov/imo/media/doc/no_fakes_act_draft_text.pdf.

⁶⁰ “Transcript: US Senate Judiciary Subcommittee Hearing on “The NO FAKES ACT,” (May. 1, 2024), <https://www.techpolicy.press/transcript-us-senate-judiciary-subcommittee-hearing-on-the-no-fakes-act/>.

⁶¹ “How Tech Giants Cut Corners to Harvest Data for A.I.,” NY Times.

⁶² Wen Wang et al., “Generative AI and Artists: Consumer Preferences for Style and Fair Compensation.”

⁶³ Kalev Leetaru, “Common Crawl And Unlocking Web Archives For Research,” Forbes (Sept. 28, 2017) <https://www.forbes.com/sites/kalevleetaru/2017/09/28/common-crawl-and-unlocking-web-archives-for-research/>.

⁶⁴ <https://www.datasetshop.com/license>.

Conclusion

Generative AI technologies are revolutionizing the art industry, offering new tools but also raising significant legal and ethical concerns. AI models are trained using expansive datasets scraped from the internet without consent, leading to loss of income and the dilution of artists' unique styles. Although synthetic data provides an alternative to organic data, its use comes with its own challenges. The legal system is struggling to keep up with these rapidly changing technologies, with courts relying on precedential cases and laws that do not address the unique nature of AI. Initiatives like the Glaze Project allow artists to mitigate unauthorized data collection for AI training, but these measures alone are insufficient to overcome the current tide and many ethical concerns. To address this evolving landscape, a shift towards opt-in data collection and licensing agreements is essential to ensuring transparency and fairness. As the industry continues to evolve, ongoing dialogue between artists, developers, and policymakers is crucial in shaping a future where technological innovation and artistic integrity can coexist.

Appendix

Content Licensing Model Types & Summary of Current Litigation

Potential to Maximize Pool of Available Training Material	License Fees	Transparency of License Terms*	Administrative Burden for Developers & Rights Holders
Direct Licensing Via 1-to-1 Negotiation			
Example: OpenAI's licensing agreement with Associated Press			
Low	Moderate to very high for most desirable content	Opaque	High
Direct Licensing Via Commercial Aggregator			
Examples: Copyright Clearance Center licensing of scientific and academic journals for AI training; Shutterstock's licensing of its image-video-music library to OpenAI, LG and Meta			
Low to moderate	Variable	Translucent to opaque	Low to moderate
Judicial Settlement Model			
Example: Proposed class settlement in Google Books lawsuit, seriously considered by book publishing industry but rejected by court			
If opt-In: Moderate; If opt-out: High	Low to moderate	Transparent to translucent	Low
Collective Licensing Model with Government/Regulatory Influence			
Example: Blanket licenses for musical works in repertories of nonprofit American Society of Composers, Authors and Publishers (ASCAP) and for-profit Broadcast Music Inc. (BMI) controlled by government consent decrees			
If opt-In: Moderate; If opt-out: High	Low to moderate	Transparent	Low
Statutory/Compulsory License			
Example: License for noninteractive digital performance of sound recordings implemented by the U.S. Copyright Act, with license fees set by Copyright Royalty Board and administered by SoundExchange			
High (no options to opt out)	Low	Transparent (but complex)	Low for creators; moderate for developers

65

⁶⁵<https://variety.com/vip/why-generative-ai-companies-will-pay-content-owners-and-licensing-models-that-will-emerge-1235944577/>.

Ongoing and Past Litigation

Case Information	Main Points	Relief Sought and Decision
<p><i>Zhang v. Google LLC</i> April 26, 2024 U.S. District Court for the Northern District of California</p>	<p>Summary: The plaintiffs are visual artists who filed a class action alleging unauthorized use of their copyrighted images to train Google’s Imagen text to image diffusion model</p> <p>Claims:</p> <ul style="list-style-type: none"> ● Direct copyright infringement against Google for the unauthorized copying of plaintiff’s work to train its AI models ● Vicarious copyright infringement against Alphabet, Google's corporate parent, for benefiting financially and continuing to benefit from the infringing activity 	<p>Relief Sought: Statutory and other damages under 17 U.S.C. § 504, destruction or other reasonable disposition of all copies defendants made or used in violation of the exclusive rights of plaintiffs and the class, and pre-judgment and post-judgment interest.</p> <p>Decision: N/A</p>
<p><i>The New York Times Company v. Microsoft Corporation</i> December 27, 2023 U.S. District Court for the Southern District of New York</p>	<p>Summary: After failed attempts to negotiate an agreement, The New York Times Company (The Times) sued Microsoft and multiple OpenAI entities, alleging copyright and related claims based on the defendants' unauthorized copying and use of copyrighted Times content to train their generative AI tools.</p> <p>Arguments:</p> <ul style="list-style-type: none"> ● Defendants are profiting off the intellectual property of others without fairly compensating the owners. ● Defendants claim their conduct is fair use because their unlicensed use of copyrighted content to train generative AI models serves a “transformative” purpose. ● This is not fair use because the AI outputs compete with and closely mimic the training data. ● Defendants knowingly removed copyright-management information from The Time’s works in violation of 17 U.S.C. § 1202(b)(1). 	<p>Relief Sought: Statutory damages, compensatory damages, restitution, and disgorgement; enjoining of infringing conduct; destruction under 17 U.S.C. § 503(b) of all GPT or other LLM models and training sets that incorporate Times works; costs, expenses, and attorney’s fees.</p> <p>Decision: N/A</p>
<p><i>Levoy v. Google LLC</i> July 11, 2023 U.S. District Court for the Northern District of California</p>	<p>Summary: Plaintiffs, a group of users of Google's products and services (including a minor), and an author allege a range of claims based on Google’s use of private, personal, and/or copyrighted materials scraped from the internet without proper consent or fair compensation to build its AI products—including the Bard chatbot, Imagen (text-to-image generative AI), MusicLM (a generative AI with text-to-music capabilities), Duet AI (generative AI used for drafting), and Gemini (multimodal machine-learning model).</p> <p>Claims:</p> <ul style="list-style-type: none"> ● A single direct copyright infringement claim (by Levoy on behalf of a copyright class) based on the unauthorized copying and use of her copyrighted book and her insights and opinions offered to various media outlets to develop and train Bard. ● Violation of California unfair competition law, the Computer Data Access and Fraud Act, negligence, invasion of privacy, intrusion upon seclusion, larceny/receipt of stolen property, conversion, trespass to chattels, intentional interference with existing contract, breach of third-party beneficiary contract, and unjust enrichment. 	<p>Relief Sought: Injunctive relief, actual and statutory damages, disgorgement of profits, punitive damages, prejudgment and post-judgment interest, and attorney’s fees and costs.</p> <p>Decision: N/A</p>

<p><i>Getty Images (U.S.), Inc. v. Stability AI, Inc.</i> February 3, 2023 U.S. District Court for the District of Delaware</p>	<p>Summary: Getty initially filed its complaint only against Stability AI, Inc., but amended the complaint to add Stability AI, Ltd. and additional jurisdictional allegations. Getty alleges defendants copied more than 12 million photographs from its copyrighted database, including associated captions and metadata, without authorization to train Stable Diffusion—defendants' AI model that provides computer images in response to text prompts.</p> <p>Claims:</p> <ul style="list-style-type: none"> • Copyright infringement based on copying and use of Getty's copyrighted images to train Stable Diffusion as well as producing derivative works. • DMCA violations based on use of a modified Getty watermark on images produced by Stable Diffusion as well as the removal of watermarks, captions, and metadata from images copied from Getty's website. • Federal trademark infringement, unfair competition, and false designation, and Delaware deceptive trade practices based on unauthorized use of Getty's trademarks with synthetic images generated by Stable Diffusion. • Dilution under federal and Delaware law based on use of Getty's famous marks on lower quality (and sometimes bizarre/grotesque) images, causing blurring or tarnishment. 	<p>Relief Sought: Permanent injunction; destruction of versions of Stable Diffusion trained using Getty images without permission; accounting to Getty for Stability's profits, gains, advantages, and value of business opportunities from the infringing acts; damages under the DMCA; profit and gain for trademark infringement and dilution and unfair and deceptive trade practices in amounts to be determined at trial, with the greater of such damages and profits trebled; statutory damages for Stability AI's willful acts of copyright infringement, its provision of false copyright management information, and its removal or alteration of Getty Images' copyright management information; statutory damages for Stability AI's willful acts of trademark infringement and unfair competition; reasonable attorney's fees; prejudgment and post-judgment interest.</p> <p>Decision: N/A</p>
<p><i>Andersen v. Stability AI Ltd.</i> January 13, 2023 U.S. District Court for the Northern District of California</p>	<p>Summary: Three artists filed a class action complaint against Stability AI, DeviantArt, and Midjourney. Plaintiffs asserted copyright and related claims based on the defendants' unauthorized use and copying of plaintiffs' works in training their AI image-generating products (Stable Diffusion, DreamUp, and Midjourney), as well as these products producing images derivative of plaintiffs' copyrighted works.</p> <p>Defendants' Motion to Dismiss: On Oct. 30, 2023, the court largely granted the defendants' motions to dismiss:</p> <ul style="list-style-type: none"> • Dismissed two of the artists' copyright claims for lack of registration. • Denied dismissal of plaintiff Andersen's direct copyright infringement claim against Stability based on copying and use of her images to train Stable Diffusion. • Dismissed direct and vicarious copyright infringement claims against DeviantArt and Midjourney. • Dismissed DMCA violations for failure to identify the type of copyright management information included in their works and which defendants were responsible. • Dismissed right of publicity claims for failure to allege any defendant used a plaintiff's name to advertise, sell, or solicit purchase of any defendant's products. • Dismissed all unfair competition claims (California, Lanham Act, common law) as either preempted or failing to allege plausible facts supporting plaintiffs' theories. 	<p>Relief Sought: Statutory damages, actual damages, disgorgement of profits, consequential damages, constructive trust over all assets created with the models, an injunction requiring DeviantArt to comply with all the terms of the DeviantArt Terms of Service, attorney's fees, and other remedies provided by law.</p> <p>Decision: Order denying motion to strike and requests for fees</p>

	<ul style="list-style-type: none"> ● Denied dismissal of declaratory relief claim ● Dismissed breach of contract claim against DeviantArt regarding its terms of service. ● Granted leave to amend all dismissed claims. <p>Amended Complaint: On Nov. 29, 2023, plaintiffs filed a first amended class action complaint adding seven artist plaintiffs and Runway as a defendant.</p> <p>Claims:</p> <ul style="list-style-type: none"> ● Direct copyright infringement against all defendants based on copying and use of plaintiffs' copyrighted works to train their models and/or that the models themselves were infringing. ● Inducement of copyright infringement against Stability and Runway based on distributing their models and promoting their ability to recreate plaintiffs' copyrighted images. ● DMCA violations against Stability, Midjourney, and Runway based on removal of watermarks, signatures, or captions during the training process and generation of copies based on original images with copyright management information altered or removed. ● Unjust enrichment against Stability, Midjourney, and Runway based on deriving profit/benefits from use of plaintiffs' works to train, develop, and promote their models. ● False endorsement against Midjourney based on its publishing a list of artists' names whose styles could be replicated to promote the capabilities of its model. ● Vicarious trade dress infringement against Midjourney based on inclusion of a clip model in its product that acts as a trade dress database and encourages its users to enter artist-name prompts and generate infringing images. ● Breach of contract against DeviantArt for violations of its terms of service entered into by several plaintiffs. 	
--	--	--

66

⁶⁶ <https://plus.lexis.com/api/permalink/a3a9fb9e-7d5b-4a72-b573-6a8e4a08971e/?context=1530671>.