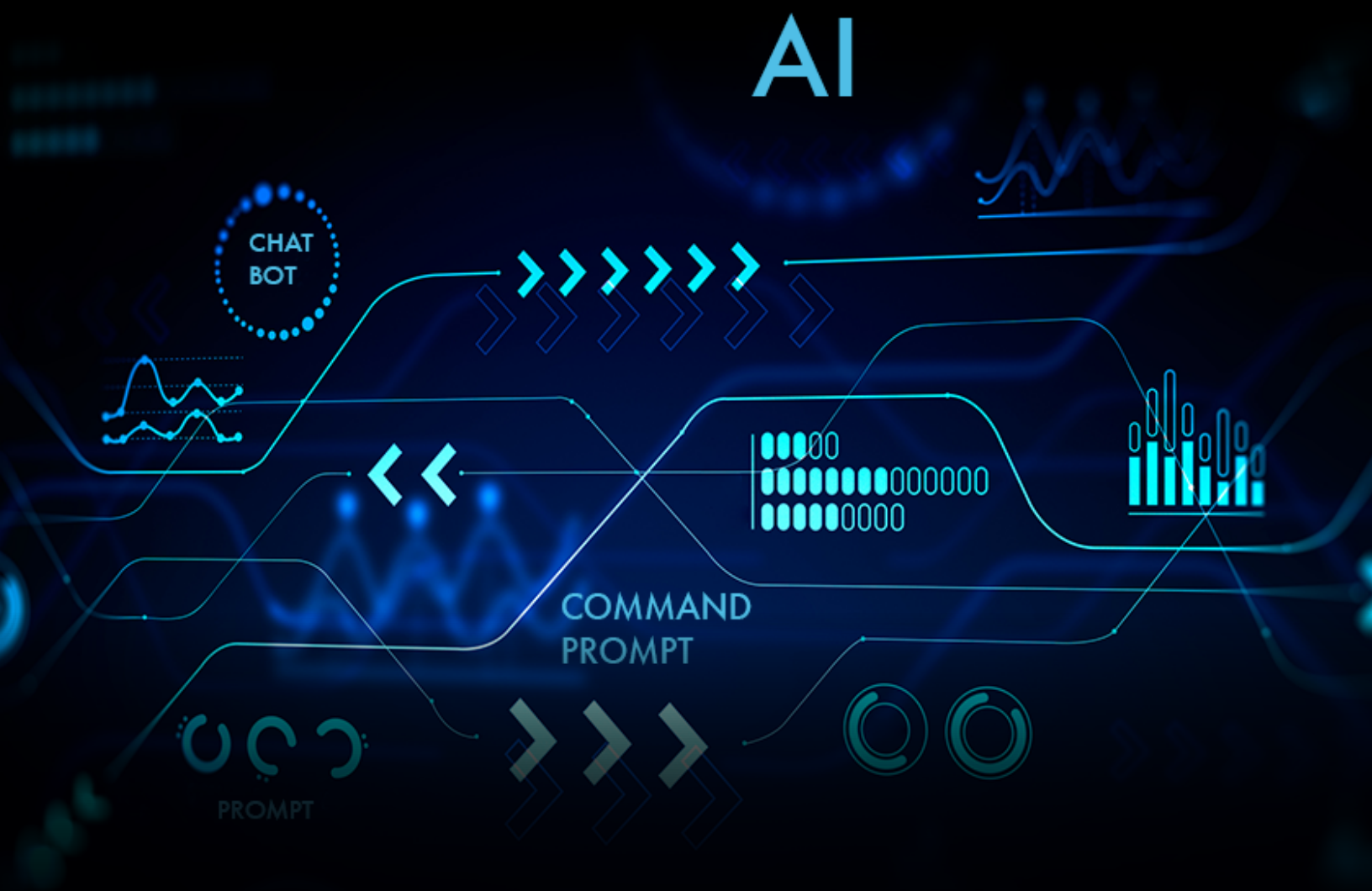
iab.

Legal Issues and Business Considerations When Using Generative AI in Digital Advertising

June 2024



Sponsored by

Lowenstein Sandler

FTI CONSULTING | TECHNOLOGY

Table of Contents

I. Introduction **5**

II. Overview of Generative Artificial Intelligence **5**

A. What Is Generative AI? **5**

B. What are Large Language Models? **6**

C. What are Weights and Biases? **8**

D. What Does It Mean To "Train" AI Models and How Are Such Models Trained? **9**

III. Uses of AI in Digital Advertising **10**

IV. Legal and Business Issues Regarding Utilizing AI to Create Content **12**

A. Can AI-Created Works Be Copyrighted? **13**

B. Can AI-Produced Works Infringe Third-Party Copyrights? **16**

C. Can AI-Produced Works Violate An Individual's Right of Publicity? **17**

D. Quality and Accuracy Risks (Hallucinations, Misinformation, and Defamation) **18**

E. Bias and Ethical Challenges **19**

F. Offensive Content; Brand Safety and Suitability **19**

G. Disclosing the Usage of AI **21**

V. Legal and Business Issues Regarding Training AI Models **23**

A. Does Ingesting Copyrighted Works Without Permission Constitute Copyright Infringement and/or Unfair Competition? **23**

1. Direct Infringement **23**
2. Vicarious Infringement **24**

B. Do Companies Ingesting the Data and Works Have a Fair Use Defense? **25**

C. Discussion of Current and Pending Copyright Infringement/Unfair Competition Actions **27**

1. Andersen, et. al. v. Stability AI, Ltd. **27**
2. Silverman, et. al. v. OpenAI and Kadrey, et. al. v. Meta Platforms **30**

D. After hiQ Labs v. LinkedIn, What Is the Law Surrounding Scraping to Train AI Models? **31**

1. Computer Fraud and Abuse Act **31**
2. Copyright Infringement **32**
3. Trespass to Chattels **33**
4. Breach of Contract **34**
5. Misappropriation **36**
6. Unjust Enrichment and Conversion **36**
7. Breach of Privacy **37**

VI. Conclusion **37**

Acknowledgements

A special thank you goes out to Lowenstein Sandler LLP and FTI Technology for sponsoring this white paper, as well as Lowenstein Sandler LLP's Matthew Savare, Bryan Sterba, and Zachary Bocian for their thought leadership and tireless drafting.

This white paper is the result of a robust and collaborative conversation between industry experts that took part in IAB's Generative A.I. working group. Thank you to Beatrice Botti, Double Verify; Matthew Epstein, Taboola; Bassam Messaike, Assistant General Counsel, Dentsu; and Nicole Sterling, Baker Hostetler for your meaningful contributions to this white paper.

All statements and opinions expressed in this white paper are those of the contributors alone and not their employers.

Disclaimer

The Interactive Advertising Bureau ("IAB") has made efforts to assure the accuracy of the material published in this whitepaper. It has been prepared for general guidance on matters of public interest only and does not constitute professional advice of any kind. You should not act upon the information contained in this publication without obtaining your own independent professional advice.

This whitepaper may also contain links and/or citations to third-party websites ("External Sites"). These links are provided solely as a convenience to you and not as an endorsement by the authors or IAB of the content on such External Sites. The content of such External Sites is developed and provided by others. You should contact the site administrator or webmaster for those External Sites if you have any concerns regarding such links or any content located on such External Sites. We are not responsible for the content of any linked External Sites and do not make any representations regarding the content or accuracy of materials on such External Sites. You should take precautions when downloading files from all websites to protect your computer from viruses and other destructive programs. If you decide to access linked External Sites, you do so at your own risk.

IAB make no representation or warranty with respect to this whitepaper, whether express or implied, including, but not limited to, any (i) implied warranty of merchantability and/or fitness for a particular purpose; and (ii) any representation or warranty concerning the accuracy or completeness of the information contained in this publication. IAB disclaims all such warranties and representations and, to the extent permitted by law, IAB does not accept or assume any liability, responsibility, or duty of care for any consequences of you or anyone else acting, or refraining to act, in reliance on the information contained in this publication or for any decision based on it.

I. Introduction

Artificial intelligence (AI), particularly generative AI, has captured the zeitgeist over the last several years and has the capacity to revolutionize countless industries, including digital advertising. Companies from across the entire ad ecosystem – from publishers and SSPs to advertisers and DSPs – are currently utilizing AI in various ways to create content, improve operational efficiencies, and optimize campaigns.

Similar to the dramatic rise in data-driven programmatic advertising, the emergence of AI is enabling market participants to create compelling content; enhance customer service and support; and improve customer segmentation, targeting, bidding, analytics, and fraud detection and prevention, all while reducing costs in a highly scalable manner.[1] However, AI also poses significant technical, operational, ethical, and legal challenges. For example, AI can be inaccurate, biased, and lacking in transparency on how content is created, or decisions are made. Moreover, in light of the enormous amount of data and content that is required to train AI models, many have raised concerns over allegations of data privacy compliance and intellectual property infringement.

This whitepaper will discuss certain legal and business issues concerning the creation, training, and usage of generative AI in digital advertising. It will provide an overview of key laws relevant to AI, including those relating to intellectual property and privacy, and a summary of select proposed laws, which if enacted, may materially impact the digital advertising industry.

II. Overview of Generative Artificial Intelligence

A. WHAT IS GENERATIVE AI?

The National Artificial Intelligence Act of 2020 ("NAIIA"), which became law on January 1, 2021,[2] defines artificial intelligence as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments."[3]

Although there are several different types of artificial intelligence,[4] this whitepaper focuses on generative AI. Generative AI is a technology that creates new content, including text, images, audio, or video, when prompted by a user.[5]

---

[1] See Richia Naidu and Martin Coulter, Reuters, From Mad Men to Machines? Big Advertiser Shift to AI, (last visited December 26, 2023), https://www.reuters.com/technology/mad-men-machines-big-advertisers-shift-ai-2023-08-18/.

[2] See National Artificial Intelligence Act of 2020 ("NAIIA"), H.R. 6216, Division E (March 12, 2020).

[3] See NAIIA, H.R. 6216, 6217.

[4] There are three different types of AI: artificial narrow AI (or weak AI), general AI (or strong AI), and super AI. Within the weak AI category, there are two functional categories: reactive machine AI (i.e., AI designed to perform a specific and narrow function, such as a recommendation engine) and limited memory AI, which utilizes large data sets to decide on a course of action to achieve a desired result. Generative AI, along with technologies such as virtual assistants, chatbots, and self-driving cars, fall within the limited memory AI category. See Understanding the Different Types of Artificial Intelligence, IBM, available at https://www.ibm.com/blog/understanding-the-different-types-of-artificial-intelligence/ (last visited February 8, 2024).

[5] See U.S. Government Accountability Office ("GAO"), GAO-23-106782 Generative AI, SCIENCE & TECH SPOTLIGHT: GENERATIVE AI (2023).

Generative AI systems create responses using algorithms that are trained on massive data sets, often publicly-available information, such as text, audio, video, and images from the internet.[6] However, generative AI systems are not cognitive and lack human judgment.[7] Using prompts (i.e., questions or descriptions entered by a user to generate and refine the results), these systems can quickly and inexpensively generate, among many other forms of materials,[8] all sorts of content for brands, agencies, and publishers, such as ad creative based on storyboards, articles or web copy, videos, images, artwork, and musical compositions.

Advanced chatbots, virtual assistants, and language translation tools are examples of mature generative AI systems in widespread use, including in digital advertising.[9] Improved computing power that can process large amounts of data for training has expanded generative AI capabilities.[10] As of early 2023, emerging generative AI systems reached more than 100 million users and attracted global attention.[11]

B. WHAT ARE LARGE LANGUAGE MODELS?

Large Language Models ("LLMs") are algorithms that recognize, summarize, translate, predict, and generate content utilizing patterns identified from very large datasets.[12] They represent a class of deep learning architectures called "transformers."[13] A transformer learns context and meaning by tracking relationships in sequential data (e.g., words in a sentence).[14] A transformer, which is made up of multiple transformer blocks known as layers,[15] has self-attention layers, feed-forward layers, and normalization layers, all working together to decipher input to predict streams of output at inference.[16] The layers can be stacked to make deeper transformers and powerful language models, and these connected layers are referred to as "neural networks."[17] Transformers were first introduced by Google in the 2017 paper "Attention Is All You Need."[18]

[6] See U.S. Government Accountability Office ("GAO"), GAO-23-106782 Generative AI, SCIENCE & TECH SPOTLIGHT: GENERATIVE AI (2023).

[7] Id.

[8] For example, Generative AI can aid complex design processes, such as designing molecules for new drugs or even creating programming code.

[9] Supra note 6.

[10] Id.

[11] Id.

[12] See NVIDA, Large Language Models Explained (last visited November 20, 2023), https://www.nvidia.com/en-us/glossary/data-science/large-language-models/.

[13] Id.

[14] Id.

[15] Id.

[16] Id.

[17] Supra note 17.

[18] See Ashish Vaswanit, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, Google Brain, Google Research, Attention is All You Need, (last visited November 20, 2023), https://arxiv.org/abs/1706.03762.

Broadly, LLMs use cases for text-based content can be divided into the following:

1. Generation (e.g., story writing, marketing content creation)
2. Summarization (e.g., paraphrasing, meeting notes summarization)
3. Translation (e.g., between languages, text-to-code)
4. Classification (e.g., toxicity classification, sentiment analysis)
5. Chatbot (e.g., open-domain Q&A, virtual assistants)[19]

Businesses across the world are leveraging LLMs to unlock a multitude of new opportunities across various industries,[20] particularly across the digital advertising ecosystem. For example, brands and agencies are using ChatGPT, copy.ai, and other LLMs to create headlines and advertising copy. Others are employing Midjourney, Dall-E 2, and similar tools to transform text-based prompts into compelling images for their ads.[21] OpenAI recently announced a new technology called Sora, which, when released to the public, will enable advertisers and other users to create high-quality videos from text descriptions.[22] Meta, Alphabet, Amazon, and other leading online advertising platforms have all debuted AI tools to help their advertisers create messages, images, and videos for their respective systems. Increasingly, marketers are using LLMs to tailor highly-personalized advertisements at scale, based on, among other things, age, geography, and interests.[23]

[19] Supra note 17.

[20] Examples outside the digital advertising space include: (i) medical researchers are training LLMs in healthcare on a corpus of data from textbooks, research papers and patient electronic health records for tasks like protein structure prediction that can uncover patterns in disease and predict outcomes; (ii) retailers are leveraging LLMs to provide stellar customer experiences through dynamic chatbots; (iii) developers are leveraging LLMs to write software and teach robots how to do physical tasks; (iv) financial advisors are using LLMs to summarize earnings calls and transcript meeting notes; and (v) marketers are training LLMs to organize customer feedback and requests into clusters or segmenting products into categories based on product descriptions. See e.g., Abraham Stern, NVIDIA, NVIDIA Expands Large Language Models to Biology, (last visited November 20, 2023), https://blogs.nvidia.com/blog/bionemo-large-language-models-drug-discovery/, Vanessa Braunstein and Nate Bradford, NVIDIA, Predict Protein Structures and Properties with Biomolecular Large Language Models, (last visited November 20, 2023), https://developer.nvidia.com/blog/predict-protein-structures-and-properties-with-biomolecular-large-language-models-2/, Virginia Jameson, SambaNova Systems, BusinessWire, New GPT AI-Powered Large Language Model for Banking Increases Financial Services Institutions' Competitiveness and Enables Accelerated Digital Transformation in Weeks, Not Years, (last visited November 20, 2023), https://www.businesswire.com/news/home/20220223005437/en/New-GPT-AI-Powered-Large-Language-Model-for-Banking-Increases-Financial-Services-Institutions-Competitiveness-and-Enables-Accelerated-Digital-Transformation-in-Weeks-Not-Years.

[21] Jonathan Vanian, How the Generative A.I. Boom Could Forever Change Online Advertising, https://www.cnbc.com/2023/07/08/how-the-generative-ai-boom-could-forever-change-online-advertising.html (last visited February 8, 2024).

[22] Ginger Adams Otis, The Wall Street Journal, New OpenAI Technology Can Create Realistic Video From a Line of Text (February 16, 2024).

[23] Vanian, supra, note 20.

As discussed in greater detail below, Large Language Models are trained using a combination of supervised and unsupervised learning.[24] With unsupervised learning, models can find previously unknown patterns in data using unlabeled datasets.[25] This also eliminates the need for extensive data labeling and redirecting when engaged in reinforced learning, which is one of the biggest challenges in building AI models.[26] Further, the significant capital investment, large datasets, technical expertise, and large-scale compute infrastructure necessary to develop and maintain LLMs make it difficult for most businesses to delve into and curate LLMs for themselves.[27]

C. WHAT ARE WEIGHTS AND BIASES?

Weights and biases are neural network parameters developed during the training process.[28] Weights and biases develop how a neural network propels data flow forward through each node of the network; this is called forward propagation, which is where input data points are each assigned a coefficient indicating how much "weight" each such data point should be given relative to the other data points when entering the next node of the network, in a forward direction, to generate an output.[29]

More specifically, weights are the relative strength of the different connections between nodes after model training, which can be likened to a human brain that has learned, for example, how to multiply numbers or speak French.[30] Further, weights help determine the importance of any given variable, with larger ones contributing more significantly to the output compared to the input.[31] All inputs are then multiplied by their respective weights and then summed.[32] Afterward, the output is passed through an activation function, which determines if that output exceeds a given threshold. If so, the activation function "fires" (or activates) the node, passing data to the next layer in the network. This results in the outputs of earlier nodes becoming the inputs of the next node, continuing to filter forward.

[24] See Isha Salian, NVIDIA, SuperVize Me: What's the Difference Between Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning?, (last visited November 20, 2023), https://blogs.nvidia.com/blog/supervised-unsupervised-learning/.

[25] See NVIDA, Large Language Models Explained (last visited November 20, 2023), https://www.nvidia.com/en-us/glossary/data-science/large-language-models/.

[26] Id.

[27] Id.

[28] See e.g., AI & Data Today, Glossary of Artificial Intelligence (AI), Machine Learning (ML), and Big Data Terms, (last visited November 20, 2023), https://www.aidatatoday.com/glossary/weight-machine-learning-term/, H2O.ai, What are Weights and Biases, (last visited November 20, 2023), https://h2o.ai/wiki/weights-and-biases/.

[29] Id; See H2O.ai, What is A Neural Network, (last visited November 20, 2023), https://h2o.ai/wiki/forward-propagation/.

[30] See Synced Review, Are Weights Really Important to Neural Networks?, (last visited November 20, 2023), https://syncedreview.com/2019/06/13/are-weights-really-important-to-neural-networks/.

[31] See IBM, What Are Neural Networks?, (last visited November 20, 2023), https://www.ibm.com/topics/neural-networks.

[32] Id.

D. WHAT DOES IT MEAN TO "TRAIN" AI MODELS AND HOW ARE SUCH MODELS TRAINED?

Machine learning is a subfield of artificial intelligence, broadly defined as the capability of a machine to imitate intelligent human behavior.[33] Machine learning starts with gathering and preparing data (e.g., numbers, photos, text, pictures of cats) to be used as training data (i.e., the information on which the machine learning model will be trained).[34] The more data, the better the program.[35]

There are three subcategories of machine learning. First, supervised machine learning models are trained with labeled data sets, which allow the models to learn and grow more accurate over time.[36] For example, an algorithm would be trained with pictures of dogs, all labeled by humans, and the machine would learn ways to identify pictures of dogs on its own. Supervised machine learning is the most common type used today.[37]

Second, in unsupervised machine learning, a program looks for patterns in unlabeled data.[38] Unsupervised machine learning can find patterns or trends for which people are not explicitly looking.[39] For example, an unsupervised machine learning program may look through online sales data and identify different types of clients making purchases.[40]

Finally, reinforcement machine learning trains machines through trial and error to take the best action by establishing a reward system, and without presenting labeled input and output pairs to the algorithm.[41] The reward system developed for the algorithm will seek to maximize accurate and correct output and will not require direct correction of flawed (or partially flawed) output, as the algorithm seeking reward maximization will avoid actions that result in a lesser or no reward. Reinforcement learning can train models to play games or train autonomous vehicles to drive by telling the machine when it made the right decision, which helps it learn over time what actions it should take.[42]

[33] See Sara Brown, MIT Sloan School of Management, Machine Learning, Explained, (last visited November 20, 2023), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.

[34] Id.

[35] Id.

[36] Id.

[37] Id.

[38] Supra note 40.

[39] Id.

[40] Id.

[41] Id.

[42] Id.

AI training is technically demanding and expensive.[43] But, years of research aided by the latest technology are helping even novice developers harness the power of original AI models to create new software.[44] However, the process of training enterprise-level AI remains incredibly difficult.[45] Data scientists may spend years creating a single new AI model and training it to perform complex tasks such as autonomous navigation, speech recognition, and language translation.[46] Although these trained AI models open a multitude of new possibilities for businesses, there are several issues business leaders must consider in using them, including understanding their limitations.[47]

III. Uses of AI in Digital Advertising

Artificial intelligence has been used in digital advertising for over a decade, and with the relatively recent emergence of generative AI, is poised to revolutionize the industry.[48] The rise of programmatic advertising, the proliferation of "big data," improvements in technology, and the dramatic decrease in compute costs have created an ideal environment for AI to propagate. Although this whitepaper is focused on generative AI, a brief description of AI's other uses in advertising is useful to show how pervasive the technology has become in a short amount of time.

Programmatic ad platforms have used artificial intelligence and machine learning for years to optimize real-time bidding.[49] Today, all ad exchanges, networks, and social media platforms use AI to a certain degree to manage real-time ad buying, selling, and placement at scale.[50] Although there are dozens of ways in which AI is being used in digital advertising, here are a few of the most pervasive and powerful ones.

First, AI is being used for audience segmentation and targeting. Although AI-driven platforms require fewer interest categories for precise targeting,[51] machine learning and AI algorithms are able to examine millions of data points about a customer to decide frequency and effectiveness when serving ads.[52]

[43] See KJ Jacoby, Performance Intensive Computing, Tech Explainer: What is AI Training?, (last visited November 20, 2023), https://www.performance-intensive-computing.com/objectives/tech-explainer-what-is-ai-training?itc=refresh.

[44] Id.

[45] Id.

[46] Id.

[47] Id.

[48] Christian Libonati and Mauro Fernandez, Globant, The Digital Advertising Revolution: How Artificial Intelligence Is Changing the Game, (last visited February 11, 2024), https://stayrelevant.globant.com/en/technology/create/ai-is-changing-the-digital-advertising-landscape/.

[49] Abid Haleem, et al., International Journal of Intelligent Networks 3, Artificial Intelligence (AI) Applications for Marketing: A Literature-Based Study (2022).

[50] Mike Kaput, Marketing Artificial Intelligence Institute, AI in Advertising: Everything You Need to Know, (last visited February 11, 2024), https://www.marketingaiinstitute.com/blog/ai-in-advertising.

[51] Libonati and Fernandez, supra, note 45.

[52] Abid Haleem, et al., supra, note 46.

Second, advertisers are employing AI to optimize their ad budgets and performance. For example, AI – without any human intervention – can utilize past campaign metrics, allocate ad dollars across various channels and audiences, and automatically adjust these decisions to reach advertising KPIs and optimize budgets.[53]

Third, platforms are using AI to generate detailed, comprehensive, and bespoke reports measuring campaign performance, which enable advertisers to gauge return on investment and adjust their strategies.[54] Equally important, these platforms enable advertisers to create algorithmic rules around campaign strategies and KPIs that can be activated or deactivated based on specific metrics to improve performance.[55] In essence, these "set it and forget it" products enable advertisers to test multiple assets and campaign strategies automatically with minimal oversight and the ability to rapidly pivot to other strategies.

Fourth, big data and artificial intelligence are helping advertisers develop deeper and more precise insights about their audiences, which may reveal certain metrics such as audiences' gender, age, and other demographic information; interests; and purchasing behavior. Such rich information enables brands to create more effective and targeted ads for each audience segment and to detect and even anticipate trends.[56]

Fifth, marketers are using AI tools to test hundreds or thousands of variations of ad creatives quickly and automatically. Brands are then using the insights gleaned from these tests to refine their ads, improve their campaigns, and increase their return on investment.[57]

Sixth, as in many other industries, marketers are utilizing AI-powered chatbots to interact with their customers in a more nuanced, personalized, and effective manner. Aside from enhancing the customer service experience, chatbots can create campaign assets, which can make the advertising process more efficient and impactful.[58]

Finally, the most publicized usage of AI in the digital advertising industry – and many other industries such as media and entertainment – is the generation of content (i.e., specific campaign assets such as photos, videos, text, or the creation of the actual advertisements). Leveraging two AI-powered technologies: natural language processing and natural language generation, these AI tools are proficient in generating "short, punchy copyrighting that often succeeds in digital advertising."[59] Increasingly, however, marketers are using generative AI to create "stunningly creative, artistic, and photo-realistic results using off-the-shelf technology."[60] In just a few years, generative AI has captured the zeitgeist and enabled advertisers to create "breathtaking visuals" for a fraction of the cost and time.[61]

---

[53] Kaput, supra, note 47.

[54] Libonati and Fernandez, supra, note 45.

[55] Id.

[56] Abid Haleem, et al., supra, note 46.

[57] Kaput, supra, note 47.

[58] Libonati and Fernandez, supra, note 45.

[59] Kaput, supra, note 47.

[60] Id.

[61] Id.

When combined with the rich and massive data sets described above and the predictive analytics fueled by these technological advancements, generative AI is quickly becoming a critical tool for the digital advertising industry. Aside from generating compelling ads, AI enables the creation of hyper-personalize ad messages and targeting to individual consumers.[62] In addition, brands can use generative AI to create multiple ad variations (e.g., different designs and creatives) and formats (e.g., ad sizes and technical specifications for various platforms) automatically and simultaneously.[63] This process removes many manual steps and helps brands stretch their marketing budgets further. Because advertisers are able to generate more content for the same or less money, they are able to test different ad creatives, refine their campaigns, and even create new ad formats, such as augmented reality ads.[64]

Despite all of these benefits, the use of AI in digital advertising comes with risks and uncertainties. For example, AI may lack the contextual understanding and empathy necessary to create content that resonates with customers. Similarly, because generative AI models are based, and therefore dependent, on the data on which they were trained, AI may lack the creativity needed to develop innovative and unconventional ads that can break through the clutter.[65] And, as discussed in greater detail below, there are a host of legal, technical, ethical, and operational risks and challenges that will need to be addressed in the coming years.

IV. Legal and Business Issues Regarding Utilizing AI to Create Content

Advertisers, agencies, and publishers – including many of the largest and most influential in the world – are increasingly using AI to create content for a fraction of traditional production costs.[66] This transformation creates myriad intellectual property, privacy, ethical and other legal issues. For instance, are AI-generated works protected by copyright? Can they infringe the intellectual property rights of third parties, such as copyrights or rights of publicity? How accurate and unbiased is AI-generated content and can inaccurate content form the basis of a defamation claim? What ethical obligations exist or should exist when utilizing AI? Does the usage of AI technology need to be disclosed, and, if so, how and under what circumstances? We explore these questions below.

[62] Id.

[63] Id.

[64] Libonati and Fernandez, supra, note 45.

[65] Id.

[66] Naidu and Coulter, supra, note 1.

A. CAN AI-CREATED WORKS BE COPYRIGHTED?

Despite a multiplicity of jurisdictional approaches to copyrighting AI-generated works, as a general matter, it has long been the position of the U.S. Copyright Office that there is no copyright protection for works created by non-humans, including machines.[67] In Thaler v. Perlmutter, U.S. District Court Judge Howell affirmed the U.S. Copyright Office's rejection of an application filed by computer scientist Stephen Thaler on behalf of his DABUS system.[68] In his opinion, Judge Howell reasoned that although copyright law has adapted with the advent of new technologies, there "has been a consistent understanding that human creativity is the sine qua non at the core of copyrightability, even as that human creativity is channeled through new tools or into new media."[69] As of October 18, 2023, the case is being appealed as it works its way up through the courts.

As to the question of whether a generative AI model can be an "author" under the Copyright Act of 1976, Judge Howell answered this question "no." Judge Howell further stated:

> To be sure, as [Thaler] points out, the critical word 'author' is not defined in the Copyright Act.[70] 'Author,' in its relevant sense, means 'one that is the source of some form of intellectual or creative work,' '[t]he creator of an artistic work; a painter, photographer, filmmaker, etc.'[71] By its plain text, the 1976 Act thus requires a copyrightable work to have an originator with the capacity for intellectual, creative, or artistic labor. Must that originator be a human being to claim copyright protection? The answer is yes.[72]

This rule begins to blur when a human and artificial intelligence collaborate to create an original work. Is the work copyrightable then? Courts have yet to answer this question, but academics have opined on the issue, the US Copyright Office Review Board has issued several decisions, and the Copyright Office has provided guidance. For example, Daniel Gervais, a professor at Vanderbilt Law School believes that "[i]f a machine and a human work together,

[67] See Ellen Glover, Brennan Whitfield, AI-Generated Content and Copyright Law: What We Know, (last visited December 22, 2023), https://builtin.com/artificial-intelligence/ai-copyright.

[68] See Thaler v. Perlmutter, No. CV 22-1564 (BAH). Litigation is currently ongoing between Thaler and the Copyright Register as to whether he can claim copyright to an AI-generated work. In his copyright registration application, Thaler claimed that the machine, not Thaler, was the author, and he purported to transfer ownership to himself under the work-made-for-hire doctrine. Naming the machine as the author is in direct contravention of Copyright Office policy. See Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190 (Mar. 16, 2023) (specifying that "[t]o qualify as a work of 'authorship' a work must be created by a human being" and that the Copyright Office "will not register works produced by a machine or mere mechanical process that operates randomly or automatically without any creative input or intervention from a human author").

[69] Thaler, No. CV 22-1564 (BAH).

[70] See Pl.'s Mem. at 24.

[71] Author, MERRIAM-WEBSTER UNABRIDGED DICTIONARY, (last visited August, 18, 2023) https://unabridged.merriam-webster.com/unabridged/author; Author, OXFORD ENGLISH DICTIONARY, (last visited August 10, 2023), https://www.oed.com/dictionary/author_n.

[72] Supra note 197.

but you can separate what each of them has done, then [the copyright] will *only* focus on the human part."[73] He opined that if, however, the human and machine's contributions are intertwined, a work's copyright eligibility hinges on how much control or influence the human author had on the machine's outputs.[74]

Another law professor at Emory University, Matthew Sag, states that artists giving instructions to AI must be sufficiently detailed to warrant copyright protection.[75] Sag further argues that: "applicants will need to do more than show that they pulled a lever on a slot machine 600 times until they got a result they were happy with."[76]

Sag's analysis, however, has not materialized. For example, in September 2022, Kris Kashtanova sought and obtained a copyright registration for a comic for which she generated images from a generative AI tool.[77] Although the Copyright Office initially approved her application, the US Copyright Office Review Board cancelled the registration and reissued a new one only in the comic as a compilation, refusing copyrights in the individual images, despite her laborious involvement in selecting and editing the specific images.[78] Kashtanova claims that she tested "hundreds or thousands of descriptive prompts" in Midjourney before landing on "as perfect a rendition of her vision as possible."[79] Despite this active level of involvement, the US Copyright Office Review Board refused her registration for the images claiming that her efforts did not meet copyright law's authorship standards. It rested its position on legislative history, language in the Constitution, and case law such as Burrow-Giles Lithographic Co. v. Sarony, which suggests that Kashtanova's involvement in ideating and executing the images falls short of authorship.[80]

For advertisers and publishers that rely on intellectual property as one of their core assets, the uncertainty surrounding the protectability of AI-generated works is a vexing problem. With no court decisions on this important matter and no amendments to the Copyright Act to address AI specifically, content creators should consider following the Copyright Office's recently-issued guidance, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence,* which states the following:

[73] Supra note 194 (emphasis added).

[74] Id.

[75] Katie Knibbs, Wired, Why This Award-Winning Piece of AI Art Can't Be Copyrighted, (last visited December 22, 2023), https://www.wired.com/story/ai-art-copyright-matthew-allen/.

[76] Id.

[77] Id.

[78] US Copyright Office Review Board, Cancellation Decision re: Zarya of the Dawn, at 8 (Feb. 21, 2023), https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.

[79] Id.

[80] See Burrow-Giles, 111 U.S. 53, 61, 58 (1884) (holding that an "author" is the one "who has actually formed the picture," and their copyrightable work is "representative[] of [their] original intellectual conceptions").

How To Submit Applications for Works Containing AI-Generated Material

Individuals who use AI technology in creating a work may claim copyright protection for their own contributions to that work. They must use the Standard Application, and in it identify the author(s) and provide a brief statement in the "Author Created" field that describes the authorship that was contributed by a human. For example, an applicant who incorporates AI-generated text into a larger textual work should claim the portions of the textual work that is human-authored. And an applicant who creatively arranges the human and non-human content within a work should fill out the "Author Created" field to claim: "Selection, coordination, and arrangement of [describe human-authored content] created by the author and [describe AI content] generated by artificial intelligence." Applicants should not list an AI technology or the company that provided it as an author or co-author simply because they used it when creating their work.

AI-generated content that is more than de minimis should be explicitly excluded from the application. This may be done in the "Limitation of the Claim" section in the "Other" field, under the "Material Excluded" heading. Applicants should provide a brief description of the AI-generated content, such as by entering "[description of content] generated by artificial intelligence." Applicants may also provide additional information in the "Note to CO" field in the Standard Application.

Applicants who are unsure of how to fill out the application may simply provide a general statement that a work contains AI-generated material. The Office will contact the applicant when the claim is reviewed and determine how to proceed. In some cases, the use of an AI tool will not raise questions about human authorship, and the Office will explain that nothing needs to be disclaimed on the application.[81]

Failure to follow this guidance may result in a rejected copyright application, which could have many negative implications. For example, without a registration, plaintiffs cannot file copyright infringement claims to protect and enforce their rights and may lose the myriad benefits of obtaining a registration prior to any act of infringement (e.g., the ability to recover statutory damages and attorneys' fees).

[81] Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16,190 (Mar. 16, 2023), https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence.

B. CAN AI-PRODUCED WORKS INFRINGE THIRD-PARTY COPYRIGHTS?

As discussed in greater detail below, there have been a number of high-profile lawsuits alleging that the process by which AI companies collect and use third-party data and content to train their models infringes copyrights. In one of the most recent examples, The New York Times sued Microsoft and OpenAI, contending that millions of its articles were used to train ChatGPT, which allegedly directly competes with The New York Times as a source of reliable information.[82] The complaint alleges that the defendants should be held accountable for "billions of dollars in statutory and actual damages" related to their "unlawful copying and use of The Times's uniquely valuable works."[83]

The New York Times is reported to have approached Microsoft and OpenAI in April of 2023 to raise concerns about the use of its intellectual property and explore "an amicable resolution," and "technological guardrails" around generative AI products.[84] But those talks did not produce a resolution.[85] Recently, a number of major publishers have struck deals to license their works with OpenAI. For example, in July 2023, The Associated Press ("AP") negotiated an agreement under which OpenAI would license AP's archive of news stories since 1985 to train its large language models,[86] and CNN, Fox, and Time are reportedly in similar negotiations.[87]

There are, however, no reported cases of an end user of a third-party AI tool being sued for copyright infringement based on the output generated in response to user prompts. Notwithstanding, on October 12, 2023, Google Cloud announced that it would indemnify its users for third-party claims arising from allegations that (1) Google's use of training data to create any of our generative models utilized by a generative AI service, infringes a third party's intellectual property right; and (2) generated output infringes a third party's intellectual property rights.[88] Google asserts that these indemnities, when taken together, "provide comprehensive coverage for our customers who may be justifiably concerned about the risks associated with this exciting new frontier of generative AI products."[89] Other generative AI providers have followed suit, with Microsoft, Adobe, Shutterstock, and IBM now offering similar indemnities.

[82] See Michael M. Grynbaum and Ryan Mac, New York Times, The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work, (last visited December 27, 2023), https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

[83] Cite Complaint instead?

[84] Id.

[85] Id.

[86] Matt O'Brien, The Associated Press, ChatGPT-Maker OpenAI Signs Deal with AP to License News Stories (last visited February 11, 2024), https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a.

[87] Reuters, OpenAI in Content Licensing Talks with CNN, Fox and Time, January 10, 2024 (last visited February 11, 2024), https://www.reuters.com/technology/openai-content-licensing-talks-with-cnn-fox-time-bloomberg-news-2024-01-11/.

[88] Google, Shared Fate: Protecting Customers With Generative AI Indemnification, October 12, 2023 (last visited February 11, 2024), https://cloud.google.com/blog/products/ai-machine-learning/protecting-customers-with-generative-ai-indemnification?utm_source=bensbites&utm_medium=newsletter&utm_campaign=daily-digest-google-da-vinci.

[89] Id.

C. CAN AI-PRODUCED WORKS VIOLATE AN INDIVIDUAL'S RIGHT OF PUBLICITY?

The right of publicity protects an individual's right to control the "commercial" use of his or her name, image, and likeness, and in some states, other aspects of his or her persona, such as voice or signature.[90] The right of publicity is governed by state law, with thirty-eight states recognizing the right of publicity (some by statute, some by common law, and some by both).[91]

Generative AI has already been used a number of times to create advertisements depicting celebrities without their permission. For example, Tom Hanks "appeared" in a video ad for a dental plan without his knowledge or consent, prompting the actor to alert his nearly ten million Instagram followers: "BEWARE!! There's a video out there promoting some dental plan with an AI version of me. I have nothing to do with it."[92] Similarly, digital reproductions of Elon Musk, Tom Cruise, and Leonardo DiCaprio have been shown in ads without their consent.[93]

Such unauthorized uses of an individual's name, image, and likeness in an ad almost certainly violate such person's right of publicity, as the purpose is commercial in nature. The analysis becomes more complicated, however, when the use is in connection with hybrid speech (i.e., a mix of commercial and non-commercial messages) or entertainment speech. For example, an anonymous TikTok user utilized generative AI to create the song, "Heart on my Sleeve," which featured voices that mimicked Drake and The Weeknd without their authorization.[94] The video of the song, which was viewed millions of times, was removed from platforms like Spotify, TikTok, and YouTube on copyright grounds, because the file contained a copyright protected producer tag.[95] Although the song featured AI-generated voices nearly identical to Drake and The Weeknd, it is unlikely their rights of publicity were violated under current law, because – unlike other soundalike cases – the use was not commercial in nature.[96] For example, in *Waits v. Frito-Lay* and *Midler v. Ford Motor Co.*, advertisements that contained the voices of singers paid to sound like famous singers, Tom Waits and Bette Midler, were found to have violated the celebrities' rights of publicity, because they did not consent to such usage.[97]

[90] Brianne Polio and Matt Savare, ANA, The No Fakes Act and the Right of Publicity in the Age of Generative AI (December 15, 2023).

[91] Id.

[92] The Guardian, Tom Hanks Says AI Version of Him Used in Dental Plan Ad Without His Consent (October 1, 2023).

[93] Patrick Coffee, The Wall Street Journal, 'Deepfakes' of Celebrities Have Begun Appearing in Ads, With or Without Their Permission (October 25, 2022).

[94] Polito and Savare, supra, note 86.

[95] Id.

[96] Id.

[97] Waits v. Frito-Lay, Inc., 978 F.2d 1093 (9th Cir. 1992); Midler v. Ford Motor Co., 849 F.2d 460 (9th Cir. 1988).

In an effort to protect individuals' rights of publicity from unauthorized recreations from generative AI, several United States Senators proposed the Nurture Originals, Foster Art and Keep Entertainment Safe Act of 2023 or the No Fakes Act. The No Fakes Act would create a uniform, federal right of publicity law with respect to uses of images, voice, and visual likeness in sound recordings and audiovisual works (not just in commercials and advertisements).[98]

As generative AI becomes more pervasive and the quality of AI-generated images, videos, and voices improves – all of which are already occurring at dizzying paces – the unauthorized uses of individuals' personas, especially those of celebrities, athletes, and other public figures, will continue to explode.

D. QUALITY AND ACCURACY RISKS (HALLUCINATIONS, MISINFORMATION, AND DEFAMATION)

Although the purveyors of generative-AI continue to make strides towards minimizing or eliminating their AI tools' likelihood of producing false or misleading information, such technologies are still prone to "hallucinations" (i.e., where large-language models produce entirely fictitious output as if it were factual information).[99] Can a hallucination result in a finding of defamation?[100] That very question is now being litigated.

In *Walters v. OpenAI*,[101] the plaintiff, a radio host, alleged that in response to a journalist's request to summarize the allegations in a complaint, ChatGPT falsely claimed that Walters was a defendant in that lawsuit and accused him defrauding and embezzling funds.[102] Walters, who was not a party in the lawsuit ChatGPT was prompted to summarize, alleged that "[e]very statement of fact in the summary pertaining to Walters is false," which constituted libel.[103]

Although the *Walters* case is still making its way through the court system, it should serve as a warning for publishers and other content creators to ensure that any (purported) facts or insights gleaned from a generative-AI tool should be fact-checked prior to dissemination in order to ensure accuracy and compliance with journalistic standards.

[98] Polito and Savare, supra, note 86. If enacted, the No Fakes Act would: (i) create a post-mortem right of publicity that would protect one's image, voice, and visual likeness for 70 years after death; (ii) prohibit the production of digital replicas or computer-generated electronic representations that are nearly indistinguishable from the actual voice or visual likeness of an individuals without their consent; (iii) prohibit the distribution, publication, or transmission of unauthorized digital replicas that one knows is unauthorized; and (iv) create civil liability for violations equal to the greater of $5,000 per violation or the damages suffered by the injured party, with punitive damages and attorneys' fees available for willful violations. Id.

[99] Clay Calvert, American Enterprise Institute, Defamation Law and Generative AI: Who Bears Responsibility for Falsities? August 22, 2023 (last visited February 17, 2024), https://www.aei.org/technology-and-innovation/defamation-law-and-generative-ai-who-bears-responsibility-for-falsities/.

[100] For an in-depth analysis as to whether AI companies can be sued for defamation based on the output of their models, see Eugene Volokh, Journal of Free Speech Law, Large Libel Models? Liability For AI Output (2023), https://www.journaloffreespeechlaw.org/volokh4.pdf.

[101] Walters v. OpenAI, LLC, 23-A-04860-2 (June 5, 2023).

[102] Id.

[103] Id.

E. BIAS AND ETHICAL CHALLENGES

Since the early days of artificial intelligence, experts have raised concerns over bias in AI. There are three principle sources of bias in AI: training data bias, algorithmic bias, and cognitive bias.[104] Because AI systems "learn" to make decisions based on the training data they are provided, bias can be introduced if the training data over or under represents a particular group or category.[105] For example, a journalist prompted an AI-system to generate images of a "technology columnist," and the system created four images of white men, whereas a prompt to a different AI-tool for a "basketball player" returned an image of a black man.[106] This source of bias has even more serious consequences when applied to non-generative AI, such as in law enforcement's use of AI-driven facial recognition tools. Algorithmic and cognitive biases, on the other hand, are caused by programming errors incorporated into the algorithms, such as developers unfairly selecting or weighting certain decision-making factors on their own conscious or unconscious biases.[107]

Many businesses in the digital advertising industry remain wary of biases that may be hard coded into generative-AI systems[108] and as such, are instituting AI governance policies to prevent the usage of generative-AI from reproducing and perpetuating human biases, including racial and gender stereotypes.[109] Advertisers are urged to provide human oversight that ensures that AI generated content is ethical and aligned to a brand's values. The Department of Commerce's National Telecommunications and Information Administration has released detailed guidance on this topic.[110]

F. OFFENSIVE CONTENT; BRAND SAFETY AND SUITABILITY:

Advertisers, agencies, and publishers have grappled for years with fake news, misinformation, and controversial and incendiary content.[111] The propagation of generative-AI has facilitated the automation and widespread dissemination of such objectionable content, especially with respect to elections, wars, and other political issues.[112]

---

[104] IBM, Shedding Light on AI Bias With Real World Examples (last visited February 17. 2024), https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/.

[105] Id.

[106] Joanna Stern, The Wall Street Journal, Ask an AI Art Generator for Any Image. The Results Are Amazing—and Terrifying (October 19, 2022).

[107] IBM, supra, note 101.

[108] Naidu and Coulter, supra, note 1.

[109] Id.

[110] NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION, DEPARTMENT OF COMMERCE, ARTIFICIAL INTELLIGENCE ACCOUNTABILITY POLICY REPORT, AI Accountability Policy Report | National Telecommunications and Information Administration (ntia.gov) (2024)

[111] AdExchanger, Fake News, Inappropriate Content And The Rise Of The Self-Policing Platform (November 15, 2017), https://www.adexchanger.com/data-driven-thinking/fake-news-inappropriate-content-rise-self-policing-platform/.

[112] Pranshu Verma, The Washington Post, The Rise of AI Fake News is Creating a 'Misinformation Superspreader' (December 17, 2023), https://www.washingtonpost.com/technology/2023/12/17/ai-fake-news-misinformation/.

Stakeholders in the digital advertising ecosystem have traditionally addressed offensive content and brand safety through the usage of contract clauses relating to editorial adjacency guidelines.[113] For example, Section II(d) of the Standard Terms and Conditions for Internet Advertising for Media Buys One Year or Less Version 3.0 states:

> **Editorial Adjacencies.** Media Company acknowledges that certain Advertisers may not want their Ads placed adjacent to content that promotes pornography, violence, or the use of firearms, contains obscene language, or falls within another category stated on the IO ("Editorial Adjacency Guidelines"). Media Company will use commercially reasonable efforts to comply with the Editorial Adjacency Guidelines with respect to Ads that appear on Media Company Properties, although Media Company will at all times retain editorial control over the Media Company Properties.... Should Ads appear in violation of the Editorial Adjacency Guidelines, Advertiser's sole and exclusive remedy is to request in writing that Media Company remove the Ads and provide makegoods or, if no makegood can be agreed upon, issue a credit to Advertiser equal to the value of such Ads, or not bill Agency for such Ads....

Large brands and agencies often enumerate a laundry list of content that they do not wish their ads to appear near. And, over the years, this list has grown dramatically (both in size and vagueness) to include such content as "misinformation"; "disinformation"; deep fakes; hate speech that incites violence or discriminates based on race, gender, nationality, ethnicity, religion, sexual orientation, sexual identity, or other characteristics; pornography or obscenity; "offensive" or "morally reprehensible" content; illegal content; malware or auto-downloads; content that violates third-party intellectual property or is defamatory; and content featuring or promoting weapons, violence, drugs, or tobacco.

Although generative-AI has the capacity to enable the inexpensive creation of such objectionable content on a massive scale, artificial intelligence also has the capability to help identify and stop the spread of such content.[114]

Utilizing AI models that have been trained on disparate public data to create campaign assets poses challenges for marketers to maintain consistency, suitability, and fidelity to their brand. As the saying goes, "garbage in, garbage out." In light of this, advertisers are turning to bespoke models, which are trained exclusively on their own assets. Such tailored models, which can be trained using as few as ten to twenty images, help ensure that the content they generate aligns with the brand identity across different teams and campaigns.[115]

---

[113] AdExchanger, supra, note 107.

[114] Naila Dharani, Jens Ludwig, and Sendhil Mullainathan, Chicago Booth Review, Can A.I. Stop Fake News?, (January 18, 2023), https://www.chicagobooth.edu/review/can-ai-stop-fake-news (noting that AI may help fight the proliferation of fake news, but acknowledging that no single algorithm can solve the problem).

[115] Chris McKay, Adobe Unveils Firefly Services and Custom Models for Enterprises, Magantive, March 26, 2024, available at https://www.maginative.com/article/adobe-unveils-firefly-services-and-custom-models-for-enterprises/.

G. DISCLOSING THE USAGE OF AI

As the quality of the output derived from generative AI continues to improve dramatically, it is often difficult to discern whether a piece of content was created, in whole or in part, using artificial intelligence. In light of this phenomenon, there have been widespread calls for greater transparency into whether and how AI is used to generate content.[116]

In the United States, there is currently no law, rule, or regulation in effect that expressly requires a disclosure that AI was used to generate content. However, there are a number of pending laws and existing laws that are relevant to AI transparency. For example, on June 5, 2023, a New York member of the United States House of Representatives introduced the AI Disclosure Act of 2023.[117] The bill, if passed, would require AI-generated content to include the following disclaimer noting the content's sources: "DISCLAIMER: this output has been generated by artificial intelligence." The disclosure requirement would apply to videos, photos, text, audio, and/or any other AI-generated material, and the Federal Trade Commission ("FTC") would police and enforce violations.[118]

Although the FTC has not yet ruled that failing to disclose the use of generative AI to create content is misleading, Section 5 of Federal Trade Commission Act of 1914 (the "FTC Act") broadly prohibits unfair or deceptive acts or practices, so it is conceivable that such failure could be found to violate the FTC Act under certain circumstances. The FTC has held that other uses of AI in advertising may be misleading under the FTC Act, cautioning:

> If you're an advertiser, you might be tempted to employ some of these tools to sell, well, just about anything. Celebrity deepfakes are already common, for example, and have been popping up in ads. We've previously warned companies that misleading consumers via doppelgängers, such as fake dating profiles, phony followers, deepfakes, or chatbots, could result – and in fact have resulted – in FTC enforcement actions.[119]

And, on June 30, 2023, the FTC issued a notice of proposed rulemaking for the Trade Regulation Rule on the Use of Consumer Reviews and Testimonials. In it, the FTC seeks to prohibit marketers from engaging in deceptive practices with respect to their product and service reviews and testimonials, including restrictions regarding the use of AI to generate false or misleading reviews or reviews on behalf of nonexistent customers.[120] In a similar vein, the SEC

[116] Reid Blackman and Beena Ammanath, Harvard Business Review, Building Transparency into AI Projects (June 20, 2022), https://hbr.org/2022/06/building-transparency-into-ai-projects.

[117] H.R.3831 — 118th Congress (2023-2024).

[118] Id.

[119] Michael Atleson, FTC, Chatbots, Deepfakes, and Voice Clones: AI Deception for Sale (March 20, 2023), last visited February 18, 2024, https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale.

[120] Federal Trade Commission, Trade Regulation Rule on the Use of Consumer Reviews and Testimonials (June 30, 2023), https://www.federalregister.gov/documents/2023/07/31/2023-15581/trade-regulation-rule-on-the-use-of-consumer-reviews-and-testimonials.

announced a series of enforcement actions on March 18, 2024 that seek to address the exaggeration of the use of AI in products or services, or what has been called "AI washing."[121] In this way, regulators have signaled a clear focus on using AI deceptively to sell products or services, whether by overexaggerating the role AI plays in the product, or using it to create "customer" reviews.

Several states have enacted laws that the undisclosed usage of AI tools could be misleading or deceptive. For example, California's Bot Disclosure Law, which became effective on July 1, 2019, prohibits the use of undeclared bots to communicate or interact with another person in California online "with the intent to mislead the other person about [the bot's] artificial identity for the purpose of knowingly deceiving the person about the content of the communication" in order to incentivize a purchase or influence a vote.[122] Similarly, the EU Artificial Intelligence Act (discussed in more detail below), which was proposed in April 2021 and was formally adopted in April 2024,[123] requires a disclosure if content was generated by AI. The EU law, much like GDPR, is a comprehensive regulatory framework, which may serve as a model for other laws, including those in the United States.

Aside from such legal and regulatory requirements, numerous social media platforms are adding labels to AI-generated content, and a user's failure to make the disclosure would be a breach of the platform's terms of service. For example, Meta plans to introduce labels for AI-generated content on Facebook, Instagram, and Threads.[124] Similarly, TikTok devotes an entire page to its policies and procedures regarding the usage of generative AI on the platform, including AI-generated content that should be labeled, that must be labeled, and that is prohibited entirely.[125]

As a best practice, advertisers and publishers should – at a minimum – clearly and conspicuously disclose AI-generated content. Sports Illustrated, which published a number of AI-generated stories under a byline of authors who do not exist, experienced swift and severe backlash and reputational harm when it failed to disclose the extent to which artificial intelligence was used to create the articles.[126]

[121] Statement of Chair Lina M. Khan Regarding the Joint Interagency Statement on AI, April 25, 2023, https://www.ftc.gov/system/files/ftc_gov/pdf/remarks-of-chair-lina-m-khan-re-joint-interagency-statement-on-ai.pdf

[122] California Business and Professions Code § 17940.

[123] In February 2024, the European Parliament committees approved the proposed Artificial Intelligence Act, sending the proposal to a full vote during Parliament's plenary session in April 2024. European Parliament, Artificial Intelligence Act: Committees Confirm Landmark Agreement (February 13, 2024), https://www.europarl.europa.eu/news/en/press-room/20240212IPR17618/artificial-intelligence-act-committees-confirm-landmark-agreement?mkt_tok=MTM4LUVaTS0wNDIAAAGRQs27S3B_qleKxFxpIZNm8YLZgSNz76qKyaTrg813AkfSUpnRGCiUgJubWW6EpEMxeZAKdmpmRMBSb7S8WtFCUk2CqfzeVT33ijNgvVhbml_m.

[124] Nick Clegg, Meta, Labeling AI-Generated Images on Facebook, Instagram and Threads (February 6, 2024), https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/.

[125] TikTok, About AI-Generated Content (last visited February 18, 2024), https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content.

[126] David Bauder, Associated Press, Sports Illustrated Found Publishing AI Generated Stories, Photos and Authors (November 29, 2023), https://www.pbs.org/newshour/economy/sports-illustrated-found-publishing-ai-generated-stories-photos-and-authors.

V. Legal and Business Issues Regarding Training AI Models

Although legal scholars are only beginning to scratch the surface of the challenges presented by AI under the current U.S. copyright law framework, several high-level issues and topics have already lead to significant debate (and litigation). As noted, in order to train their AI models, companies must ingest massive amounts of data and works, which often include copyrighted works from third-party publisher sites. Some have argued that trained generative AI models when prompted, after ingesting all of this presumably unlicensed copyrighted material, could be indirectly "copying" millions of human beings' works without permission.

A. DOES INGESTING COPYRIGHTED WORKS WITHOUT PERMISSION CONSTITUTE COPYRIGHT INFRINGEMENT AND/OR UNFAIR COMPETITION?

1. Direct Infringement

To establish a prima facie case of copyright infringement, a copyright owner must prove "(1) ownership of a valid copyright, and (2) copying [by the defendant] of constituent elements of the work that are original."[127] While establishing originality varies by jurisdictions outside the U.S., these elements are potentially implicated by generative AI's "scraping" the internet for text and other sources of training data, the copying that occurs during the training process, and the generation of the output when prompted with tasks. Generative AI model developers have themselves acknowledged that the training process relies on "large, publicly available datasets that include copyrighted works" and that this process "involve[d] first making *copies* of the data to be analyzed."[128] There are currently multiple lawsuits[129] centered around whether this ingestion of data during the training process is considered infringement; or, more importantly, whether such infringement would fall under the "fair use" defense.

[127] See e.g., Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991); Kaplan v. Stock Mkt. Photo Agency, Inc., 133 F. Supp. 2d 317 (S.D.N.Y. 2001); Syrus v. Bennett, 455 F. App'x 806 (10th Cir. 2011).

[128] See U.S. Patent and Trademark Office ("USPTO"), USPTO Docket No. PTO-C-2019-0038 Comment of OpenAI, LP Addressing Question 3 OpenAI RFC-84-FR-58141, Before the United States Patent and Trademark Office Department of Commerce Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation (2019).

[129] See, e.g., The New York Times Co. v. Microsoft et al. (December 2023) (claiming "[b]y building training datasets containing millions of copies of Times Works, including by scraping copyrighted Times Works from The Times's websites and reproducing such works from third-party datasets, the OpenAI Defendants have directly infringed The Times's exclusive rights in its copyrighted works); Andersen v. Stability AI Ltd. (January 2023) (alleging that defendants "downloaded or otherwise acquired copies of billions of copyrighted images without permission" to create Stable Diffusion, an "AI-based image generation product"; Silverman v. OpenAI (July 2023) (alleging that OpenAI's ChatGPT was trained on illegally-acquired datasets containing the plaintiffs' works; and Kadrey v. Meta Platforms (July 2023) (alleging that Meta's LLaMA was trained on illegally-acquired datasets containing the plaintiffs' works).

2. Vicarious Infringement

Additionally, a critical issue involves whom should be liable if the generative AI models do infringe upon copyright owners' works. Even if an end user of the AI platform could be directly liable for infringement, the company overseeing and operating the generative AI model may face liability under the doctrine of "vicarious infringement."[130] This doctrine applies to defendants who have "the right and ability to supervise the infringing activity" and "a direct financial interest in such activities."[131] This is the general tort doctrine of respondeat superior applied in cases of direct copyright infringement.[132] It is important to note that vicarious infringement will not be established, unless direct infringement is first established. Additionally, the operators of the generative AI models may benefit from the "substantial non-infringing usage" doctrine, which relieves technology creators from liability for downstream infringement if the created technology was not intentionally designed for such infringement and significant "non-infringing" uses are plausible.[133]

Even in the absence of an employer-employee relationship, a company could still be vicariously liable for infringement if the company has the right and ability to supervise the infringing activity and has a direct financial interest in such activity.[134] In Shapiro v. H.L. Green Co., the defendant department store was held liable for the infringing sale of pirated records manufactured and sold by its retailing concessionaire.[135] Under the terms of the concession agreement, the store retained the ultimate right of supervision over the conduct of the concessionaire and received as rental a percentage of the concessionaire's gross sales.[136] The Second Circuit attached no special significance to the technical classification of the store-concessionaire relationship.[137] Rather it found the policies of the copyright law would be best effectuated if the store were held liable, even in the absence of actual knowledge that the copyright was being impaired, for its failure to police the conduct of the primary infringer.[138] Thus, where an operator of an AI model has the right and ability to supervise its models and their usage by customers, and they have a direct financial interest in such models' success, it is conceivable that they may be held vicariously liable for the copyright infringement the models commit based on prompts submitted by third parties.

[130] See Gershwin Publishing Corp. v. Columbia Artists Man., Inc., 443 F. 2d 1159 (2nd Cir. 1971).

[131] See Gershwin Publishing Corp., 443 F. 2d 1159 (2nd Cir. 1971).

[132] See M. Witmark & Sons v. Calloway, 22 F.2d 412, 414 (E.D. Tenn. 1927).

[133] Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984).

[134] See Shapiro, Bernstein & Co. v. H. L. Green Co., 316 F.2d 304 (2 Cir. 1963).

[135] See Shapiro, 316 F.2d 304.

[136] Id.

[137] Id.

[138] Id.

B. DO COMPANIES INGESTING THE DATA AND WORKS HAVE A FAIR USE DEFENSE?

AI model operators argue that the model training process constitutes "fair use" and does not infringe the copyright owners' works.[139] Fair use, which is an affirmative defense to copyright infringement, is based on four statutory factors under 17 U.S.C. § 107:

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.[140]

While this defense may not be available in some international jurisdictions, courts evaluate fair use claims on a case-by-case basis, and the outcome of any given case depends on a fact-specific inquiry.[141] This means there is no formula to ensure that a predetermined percentage or amount of a work – or a specific number of words, lines, pages, or copies – may be used without permission.[142]

When AI algorithms or models "learn" their functions by ingesting copyrighted works, reproductions of those works are made in the process, as the works are digitized and/or "read" by the algorithms.[143] Some mass digitization scenarios may be a fair use while others may constitute infringement.[144] In Authors Guild, the Second Circuit found that the search and accessibility use of digitized books constituted "fair use" and that Google's unauthorized digitizing of tens of millions of copyright-protected books, its creation of a search functionality, as well as its display of snippets from those books, were each not infringing uses.[145]

[139] For example, in response to the New York Times' suit, OpenAI responded that "training AI models using publicly available internet materials is fair use, as supported by long-standing and widely accepted precedents." See also Congressional Research Service, Generative Artificial Intelligence and Copyright Law, (last visited February 9, 2024), https://crsreports.congress.gov/product/pdf/LSB/LSB10922; See USPTO Docket No. PTO-C-2019-0038 Comment of OpenAI, LP Addressing Question 3 OpenAI RFC-84-FR-58141. For support, OpenAI cites the Authors Guild case.

[140] 17 U.S.C. § 107.

[141] See USPTO Docket No. PTO-C-2019-0038 Comment of OpenAI, LP Addressing Question 3 OpenAI RFC-84-FR-58141.

[142] Id.

[143] Id; See also Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014).

[144] See Authors Guild, Inc., 755 F.3d 87.

[145] Id.

Similarly, AI model operators have argued the purpose of model training is "transformative" as opposed to "expressive" (or derivative) because the training process creates "a useful generative AI system."[146] Whether the secondary use "transforms" the original in some way, or if it "merely supersede[s]" the original is a major hurdle "fair use" proponents must pass.[147] Since 1994, when the Supreme Court adopted "transformativeness" as part of the inquiry about the purpose and character of the secondary use in Campbell, this question has grown increasingly important.[148] If one can show that their secondary use transforms the original in some way, it is much more likely to be "fair use" then otherwise.[149] This leaves much to be determined as to whether generative AI algorithms truly "transform" the works they copy when they scrape.

Regarding the fourth fair use factor, some generative AI applications raised concerns among various rights holders that training AI on copyright works will allow the applications to generate similar derivative works that compete with the original copyrighted work.[150] For example, an AI-generated song called "Heart on My Sleeve," made to sound like the artists Drake and The Weeknd, was heard millions of times on streaming services.[151] Universal Music Group, who has deals with both artists, argued that AI companies violate copyright by using artists' songs in training data.[152]

Plaintiffs have filed multiple lawsuits claiming the training process for AI programs infringed their copyrights in written and visual works. These include lawsuits by the Authors Guild and authors Paul Tremblay, Michael Chabon, Sarah Silverman, and others against OpenAI; separate lawsuits by Michael Chabon, Sarah Silverman, and others against Meta Platforms; proposed class action lawsuits against Alphabet Inc. and Stability AI and Midjourney; and a lawsuit by Getty Images against Stability AI.[153] The Getty Images complaint alleges that "Stability AI has copied at least 12 million copyrighted images from Getty Images' websites . . . in order to train its Stable Diffusion model."[154] Plaintiffs in this lawsuit dispute any characterization of fair use, arguing that Stable Diffusion is a commercial product, weighing against fair use under the first statutory factor, and that the program undermines the market for the original works, weighing against fair use under the fourth factor.[155]

[146] See Congressional Research Service, Generative Artificial Intelligence and Copyright Law, (last visited November 21, 2023), https://crsreports.congress.gov/product/pdf/LSB/LSB10922; See USPTO Docket No. PTO-C-2019-0038 Comment of OpenAI, LP Addressing Question 3 OpenAI RFC-84-FR-58141.

[147] See Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 575 (1994).

[148] Campbell, 510 U.S. 569, 575.

[149] Id.

[150] See Congressional Research Service, Generative Artificial Intelligence and Copyright Law, (last visited November 21, 2023), https://crsreports.congress.gov/product/pdf/LSB/LSB10922.

[151] See Mia Soto, The Verge, Drake's AI Clone Is Here – and Drake Might Not Be Able to Stop Him, (last visited November 21, 2023), https://www.theverge.com/2023/5/1/23703087/ai-drake-the-weeknd-music-copyright-legal-battle-right-of-publicity.

[152] Id.

[153] See e.g., Silverman, et al. v. OpenAI, Inc. (3:23-cv-03416); Chabon v. Meta Platforms Inc. (3:23-cv-04663); Andersen v. Stability AI Ltd. (3:23-cv-00201); Getty Images (US), Inc. v. Stability AI, Inc. (1:23-cv-00135).

[154] See Complaint, Getty Images (US), Inc. v. Stability AI, Inc. (1:23-cv-00135).

[155] Id.

While these cases make their way through the courts, it is not clear when we will know the viability of a fair use defense in connection with AI models' scraping of internet sources for training data. A brief description of the principal pending cases is warranted.

C. DISCUSSION OF CURRENT AND PENDING COPYRIGHT INFRINGEMENT/ UNFAIR COMPETITION ACTIONS

1. Andersen, et. al. v. Stability AI, Ltd.

Plaintiff, Sarah Andersen, on behalf of a class of similarly-situated persons, brought a class action against Defendants Stability AI, Midjourney, Inc., and DeviantArt, Inc. for direct and vicarious copyright infringement as well as violations of plaintiffs' rights of publicity and violations of unfair competition law.[156] Andersen is a full-time cartoonist and illustrator, who owns the copyright interest in over two hundred works that the defendant, Stability AI, used as training data for its AI model.[157] The Andersen complaint alleges that by "scraping" the internet, Stability AI "thereby copied over five billion images from websites as the 'Training Images' used to train Stable Diffusion. Further, the complaint states that Stability did not seek consent from either the creators of the [images] (including Andersen) or the websites that hosted them from which they were scraped."[158]

Stable Diffusion[159] is Stability's AI image product that generates output images when prompted based on what Stable Diffusion learned from those images it scraped.[160] Plaintiffs allege that Stability AI directly infringed their rights, under 17 U.S.C. Section 106 of the Copyright Act, which provides the owners of a copyright the following exclusive rights:

[156] See Complaint page 1, Andersen v. Stability AI Ltd. (3:23-cv-00201).

[157] Id. at page 6.

[158] Id. at page 13.

[159] The "diffusion" in the name Stable Diffusion comes from a software technique that operates in two phases: (1) take an image and progressively add more noise to it in a series of steps until it is essentially random fluctuations perceived by humans as chaotic and unstructured; and (2) the program then runs the sequence of noise backwards, in reverse order, progressively removing the noise ("denoising") the date until the original image is reconstructed. Id.

[160] Id. at page 15.

1. to reproduce the copyrighted work in copies or phonorecords;
2. to prepare derivative works based upon the copyrighted work;
3. to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending;
4. in the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works, to perform the copyrighted work publicly;
5. in the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including the individual images of a motion picture or other audiovisual work, to display the copyrighted work publicly; and
6. in the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission.[161]

Plaintiffs claim that Stability AI directly infringed on their rights by:

1. reproducing one or more of the original works in violation of Section 106(1);
2. preparing derivative works based upon the original works in violation of Section 106(2);
3. distributing copies of the original works to the public in violation of Section 106(3);
4. performing one or more of the original works publicly in violation of Section 106(4); and/or
5. displaying one or more of the original works publicly in violation of Section 106(5).[162]

[161] 17 U.S.C. Section 106(1-6).

[162] See Complaint at page 31, Andersen v. Stability AI Ltd. (3:23-cv-00201).

The complaint also alleges that Stability AI is vicariously liable for the alleged copyright infringement because, as stated above, Stability AI had "the right and ability to supervise the infringing activity" and "a direct financial interest in" Stable Diffusion's activities.[163]

Plaintiffs do not rely solely on the claim that Stable Diffusion outputs copyrighted images.[164] They make an additional argument that the images produced by Stability AI are derivative, because they are "generated exclusively from a combination of the conditioning data and the latent images, all of which are copies of copyrighted images."[165]

It is likely factually incorrect that "all" of the images Stability AI used to train its AI model were copyrighted images, as plaintiffs allege.[166] As stated in the complaint, plaintiffs claimed Stability AI's training data included "countless copyrighted images."[167] Although this allegation is likely true, it reveals a conundrum: how can copyright infringement be found with respect to any given output image if it is possible that the AI model could have produced such image without use of the allegedly infringed copyrighted work (e.g., after being trained on a smaller set of training data that excluded the copyrighted work)?[168]

In response to the complaint, defendants filed motions to dismiss, and the judge dismissed all but one claim for direct copyright infringement by Stability AI on behalf of just one plaintiff.[169] The judge did provide plaintiffs with leave to re-plead their claims, but cautioned that they must allege more specifically how each defendant was involved in the claimed infringement.[170] For example, discussing Midjourney, the judge reasoned: "Plaintiffs need to clarify their theory against Midjourney – is it based on Midjourney's use of Stable Diffusion, on Midjourney's own independent use of training images to train the Midjourney product, or both?"[171] The judge also indicated that plaintiffs will need to show that the images Stable Diffusion outputted were "substantially similar," to plaintiffs' original works to succeed on their copyright claims, absent a showing of direct copying. *Id.* In addition, for certain claims, plaintiffs must show additional facts, where and if possible, to establish that the outputted images were so "similar to plaintiff[s'] styles or artistic identities" so as to be possible construed as "fakes."[172]

---

[163] See, e.g., Complaint at page 32-32, Andersen v. Stability AI Ltd. (3:23-cv-00201); Publishing Corp. v. Columbia Artists Man., Inc., 443 F. 2d 1159.

[164] See Gabriel Karger, Columbia Science and Technology Law Review, AI-Generated Images: The First Lawsuit, (last visited November 22, 2023), https://journals.library.columbia.edu/index.php/stlr/blog/view/479#_ftn9.

[165] Id.

[166] Id. at footnote 9.

[167] Complaint, Andersen v. Stability AI Ltd. (3:23-cv-00201).

[168] Id.

[169] See Lauren Leipold, Grayson Moronta, Owen Wolfe, JD Supra, Some Stability for AI Defendants: Judge Dismisses All But One Claim in Andersen et. al., v. Stability AI LTD,., et. al., (last visited November 22, 2023), https://www.jdsupra.com/legalnews/some-stability-for-ai-defendants-judge-4703878/#:~:text=As%20we%20previously%20wrote%2C%20the,licenses%20to%20train%20the%20programs.

[170] Id.

[171] Id.

[172] Id.

This decision provides some clarity to plaintiffs asserting AI-related copyright infringement claims and will likely have an impact on other cases, including *Silverman v. OpenAI*.[173] The "substantial similarity" issue was raised in *Silverman v. OpenAI* as well.[174] In Silverman, defendants have already submitted the Andersen judge's decision to the judge overseeing the Silverman case, thus, claiming that the decision supports their motions to dismiss Silverman's and her co-plaintiffs' claims.[175]

2. Silverman, et. al. v. OpenAI and Kadrey, et. al. v. Meta Platforms

In *Silverman, et. al. v. OpenAI* and *Kadrey v. Meta Platforms,* plaintiffs allege that OpenAI's ChatGPT and Meta's LLaMA models were trained on datasets that included their copyrighted works, such as Silverman's book, *The Bedwetter,* without permission.[176] Other plaintiffs in each of these cases include Christopher Golden, a writer, and Richard Kadrey, also a writer, who allege their copyrighted books were infringed in the same manner.[177] The complaints allege that plaintiffs never authorized OpenAI or Meta to make copies or derivative works of their copyrighted works, publicly display such copies or derivative works, or distribute such copies or derivative works in connection with the training processes for ChatGPT or LLaMA or any subsequent use thereof and that such conduct violated the Copyright Act.[178]

The complaints also allege virtually the same other causes of action (e.g., vicarious copyright infringement and unfair competition) that Andersen alleged against Stability AI as noted above. The difference here being the alleged infringed copyrights were books and text, rather than images and illustrations.

The Silverman complaint also specifically alleges that "[w]hen ChatGPT was prompted to summarize books written by each of the Plaintiffs, it generated very accurate summaries." Although small details were incorrect, "the rest of the summaries [were] accurate, which mean[t] that ChatGPT retain[ed] knowledge of particular works in the training dataset and [was] able to output similar textual content."[179] The plaintiffs also claim that OpenAI and Meta intentionally removed copyright-management information, such as copyright notices and titles, in creating derivative works of the books.[180]

[173] Id.

[174] Supra note 108.

[175] Id.

[176] See Complaint at page 9, Silverman, et al. v. OpenAI, Inc. (3:23-cv-03416).

[177] Id. at pages 9 and 10.

[178] See Complaint at page 11, Silverman, et al. v. OpenAI, Inc. (3:23-cv-03416); See Complaint at page 3 – 6, Kadrey, et al. v. Meta Platforms, Inc. (3:23-cv-03417); 17 U.S.C. Section 106.

[179] See Complaint at page 9 – 13, Silverman, et al. v. OpenAI, Inc. (3:23-cv-03416).

[180] See Complaint at page 9 – 13, Silverman, et al. v. OpenAI, Inc. (3:23-cv-03416); See Complaint at page 3 – 6, Kadrey, et a.. v. Meta Platforms, Inc. (3:23-cv-03417).

On February 12, 2024, Judge Araceli Martínez-Olguín dismissed the Silverman and Tremblay complaints against OpenAI with respect to vicarious infringement, violation of the Digital Millennium Copyright Act, and negligence with leave to amend. The judge also dismissed with prejudice the unjust enrichment claim, but permitted the unfair competition claim to proceed. The plaintiffs' claim of direct infringement was not part of OpenAI's motion to dismiss, so that claim will proceed as well.[181]

D. AFTER HIQ LABS V. LINKEDIN, WHAT IS THE LAW SURROUNDING SCRAPING TO TRAIN AI MODELS?

In *hiQ Labs v. LinkedIn*, the United States Ninth Circuit Court affirmed the district court's granting of a preliminary injunction that prevented LinkedIn from denying HiQ Labs from accessing and scarping LinkedIn's publicly-available LinkedIn member profiles.[182] hiQ was a small data analytics company that used its own automated bots to scrape information from public LinkedIn profiles.[183] Scraping, as noted above, is a process of extracting data from a website and copying it into a structured format, allowing for data manipulation or analysis; scraping can be done manually, but is typically done by a web "bot."[184]

In finding for hiQ, the Ninth Circuit found that "hiQ established a likelihood of irreparable harm because the survival of its business was threatened [by denying it access to the data]."[185] As discussed below, the Court also made findings under the Computer Fraud and Abuse Act ("CFAA") and affirmed that the public interest favored the granting of the preliminary injunction in favor of hiQ.[186]

The following subsections delve into whether *hiQ* affects the bringing of causes of actions with respect to scraping under the following enumerated legal regimes.

1. Computer Fraud and Abuse Act

In response to a LinkedIn cease-and-desist letter sent in May 2017, hiQ filed for a preliminary injunction partially based on the claim that LinkedIn could not lawfully invoke the CFAA against hiQ.[187] The CFAA aims to address computer hacking and prohibits accessing a "protected computer" without authorization.[188] In 2019, the Ninth Circuit

[181] Order, Tremblay et al. v. OpenAI, Inc., Case Nos. 23-cv-03223-AMO and 23-cv-03416-AMO (February 12, 2024), https://www.law360.com/dockets/download/65cb73da20298f0061daa8c6?doc_url=https%3A%2F%2Fecf.cand.uscourts.gov%2Fdoc1%2F035124068003&label=Case+Filing.

[182] See HiQ Labs v. LinkedIn, 938 F.3d 985 (Sept. 9, 2019).

[183] See hiQ Labs, 938 F.3d 985, 996.

[184] See Carrie Grosvenor, EasyTechJunkie, What Is a Screen Scraper?, (last visited November 29, 2023), http://www.wisegeek.com/what-is-a-screen-scraper.html.

[185] Id.

[186] Id.

[187] See Katrina Slack, Morgan Lewis Bockius, Data Scraping Deemed Legal in Certain Circumstances, (last visited November 30, 2023), https://www.morganlewis.com/blogs/sourcingatmorganlewis/2022/04/data-scraping-deemed-legal-in-certain-circumstances.

[188] Id.

held that hiQ's activities were not "without authorization," because the data scraped was publicly available.[189] In 2022, on remand from the United States Supreme Court, the Ninth Circuit once again held that the concept of "without authorization" does not apply to publicly-available data.[190]

Thus, as publishers are recognizing, violations of terms of service that prohibit scraping will likely not enable publishers to state a cause of action under the CFAA with respect to publicly-available data.[191] Despite the common misperception that the *hiQ* decision permits all forms of data scraping, aggrieved publishers may have recourse other than under the CFAA. As the Ninth Circuit noted: "[e]ntities that view themselves as victims of data scraping are not without resort, even if the CFAA does not apply: state law trespass to chattels claims may still be available. And other causes of action, such as copyright infringement, misappropriation, unjust enrichment, conversion, breach of contract, or breach of privacy, may also lie.[192]

2. Copyright Infringement

Although data – as a mere collection of facts – is not protectible under copyright in the United States,[193] databases, under certain conditions, can be copyrighted as compilations. Similarly, web pages are protectible under copyright law. Accordingly, publishers may have recourse under copyright law with respect to the scraping of content from their websites. Two cases from the U.S. District Court for the Northern District of California have specifically addressed this issue: *Facebook Inc. v. Power Ventures, Inc.* and *Craigslist Inc., v. 3Taps Inc.*

In *Facebook Inc. v. Power Ventures Inc.*, defendant Power.com provided a service that aggregated and displayed social networking and email accounts on a single portal.[194] Users provided Power.com with login credentials to various social media websites, and Power.com's software would use these credentials to scrape data from websites like Facebook.[195] Facebook alleged that this service violated registered copyrights it maintained on its webpages, and Power.com moved to dismiss the copyright claims.[196] The district court denied the motion, holding that "if Defendants first have to make a copy of a user's entire Facebook profile page in order to collect that user content, such action may violate Facebook's proprietary rights."[197]

[189] Id.

[190] See Katrina Slack, Morgan Lewis Bockius, Data Scraping Deemed Legal in Certain Circumstances, (last visited November 30, 2023), https://www.morganlewis.com/blogs/sourcingatmorganlewis/2022/04/data-scraping-deemed-legal-in-certain-circumstances.

[191] Id.

[192] See, e.g., Associated Press v. Meltwater U.S. Holdings, Inc., 931 F. Supp. 2d 537, 561 (S.D.N.Y. 2013) (holding that a software company's conduct in scraping and aggregating copyrighted news articles was not protected by fair use)." See hiQ, 938 F.3d 985, 1047 (emphasis added).

[193] See Feist Publications, Inc. v. Rural Telephone Service Co., 499 U. S. 340 (1991).

[194] No. C 08-5780 JF (RS), 91 U.S.P.Q.2d 1430 (N.D. Cal. May 11, 2009).

[195] Id.

[196] Id.

[197] Id.

In *Craigslist, Inc. v. 3Taps Inc.,* defendants aggregated and republished ads from Craigslist's website.[198] These ads are user-generated and include user data, but Craigslist's terms of use grant Craigslist an "ownership interest to assert" copyrights that encompass the ads and data.[199] Craigslist maintained copyrights on compilations of these ads, organized by, for example, geographic location and category.[200] The district court denied defendants' motion to dismiss the copyright claims in part, finding that Craigslist had a copyright on the organization of the ads.[201]

Facebook and *Craigslist* show that website owners may be able to assert copyright claims against web scraping of user data in at least two circumstances: where more than just the user data is scraped, and where the website owner has an exclusive license to, or ownership of, the scraped data.[202] Other copyright cases relating to web scraping outside the user data context also provide guidance:

1. *Associated Press v. Meltwater U.S. Holdings, Inc.* (holding that scraping is unlikely to be a fair use);
2. *DHI Group, Inc. v. Kent* (holding that registered copyrights are presumed valid and prevail over web scraping when the scraping involves distinctive webpages); and
3. *Compulife Software, Inc. v. Newman* (holding that preparing to launch scraping software may itself be copyright infringement where the creation of such software involved reliance on copyrighted elements scraped from target online database).[203]

3. Trespass to Chattels

In footnote 21 of its 2022 *hiQ* decision, the Ninth Circuit stated in dicta that "LinkedIn's cease-and-desist letter also asserted a state common law claim of trespass to chattels.[204] Although we do not decide the question, it may be that web scraping exceeding the scope of the website owner's consent gives rise to a common law tort claim for trespass to chattels, at least when it causes demonstrable harm."[205] To bolster this premise, the Ninth Circuit cited various precedents. For example, in *eBay, Inc. v. Bidder's Edge, Inc.*,[206] the court found that eBay had established a

[198] Craigslist, Inc. v. 3Taps Inc., 942 F. Supp. 2d 962, 971 (N.D. Cal. 2013).

[199] Id.

[200] Id.

[201] Id.

[202] See Wendy Ray, David Nathanial Tan, and Jacob Nagy, Morrison Foerster, Bloomberg BNA, The Copyright Defense to Web Scraping, (last visited December 4, 2023), https://www.bloomberglaw.com/document/X10SUETC000000.

[203] Id.

[204] See hiQ, 938 F.3d 985, 1046, n. 21.

[205] Id.

[206] Id.

likelihood of success on the merits of its trespass to chattels claim against the auction-aggregating site, Bidder's Edge. The court held that although eBay's site was "publicly available," its servers were private property, and Bidder's Edge's data scraping – even though it did not cause physical harm – exceeded eBay's terms of service.[207]

Similarly, in *Register.com, Inc. v. Verio, Inc.* the Second Circuit held that a company that scraped a competitor's website to obtain data for marketing purposes likely committed trespass to chattels, because scraping could – although it did not yet – cause physical harm to plaintiff's computer servers.[208] And, in *Sw. Airlines Co. v. FareChase, Inc.*, the Northern District of Texas held that the use of a scraper to glean flight information was unauthorized as it interfered with Southwest's use and possession of its site, even if the scraping did not cause physical harm or deprivation.[209]

Accordingly, publishers may have a cause of action for trespass when their websites are scraped.

4. Breach of Contract

Likewise, publishers may be able to state a claim for breach of contract in connection with unauthorized web scraping. Although often overlooked, the Ninth Circuit did rule that hiQ breached its user agreement with LinkedIn.[210] LinkedIn alleged a breach of contract claim on two issues: (1) hiQ's automated web scraping of LinkedIn profiles; and (2) hiQ's engagement of independent contractors or "turkers" to log into LinkedIn with fake profiles to conduct quality assurance for hiQ. The LinkedIn user agreement stated as follows:

> Users agree not to '[s]crape or copy profiles and information of others through any means (including crawlers, browser plugins and add-ons, and any other technology or manual work),' '[c]opy or use the information, content or data on LinkedIn in connection with a competitive service (as determined by LinkedIn),' '[u]se manual or automated software, devices, scripts robots, other means or processes to access, 'scrape,' 'crawl' or 'spider' the Services or any related data or information,' or '[u]se bots or other automated methods to access the Services.'[211]

[207] See eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1070 (N.D. Cal. 2000).

[208] See Register.com, Inc. v. Verio, Inc., 356 F. 3d 393, 437-438 (2d Cir. 2004).

[209] See Sw. Airlines Co. v. FareChase, Inc., 318 F. Supp. 2d 435, 442 (N.D. Tex. 2004).

[210] See hiQ, 938 F.3d 985.

[211] See hiQ, 938 F.3d 985, 996 n. 6.

Although the Ninth Circuit concluded that hiQ did breach its user agreement with LinkedIn, it nonetheless denied summary judgment to LinkedIn so that a jury could consider hiQ's affirmative defenses.[212] The Court specifically held that hiQ was subject to LinkedIn's user agreement, which expressly prohibited web scraping without LinkedIn's permission and the creation of fake profiles. The Court also found that "[r]egardless of whether the turkers scraped LinkedIn's site, they breached the [u]ser [a]greement's prohibition on creating false identities."[213]

Meta Platforms, Inc. v. Bright Data Ltd.[214] is the most recent case to analyze a breach of contract claim in connection with web scraping. In *Bright Data,* Meta alleged, among other claims, that Bright Data's web scraping activities and resale of information obtained from publicly-available portions of Facebook and Instagram violated the platforms' terms. On January 23, 2024, after thoroughly analyzing the terms and applying traditional principles of contract interpretation, Judge Edward M. Chen granted Bright Data's motion for summary judgment, holding that Bright Data did not "use" Facebook and Instagram when it engaged in public logged-off scraping.[215] And, since Bright Data did not engage in scraping when it was logged into (and thus "using") the platforms, there was no cause of action for breach of contract: "When an entity does not utilize that access to, e.g., scrape public data, it does not abuse that access; it stands in the same shoes as a visitor to whom the Terms cannot apply as a matter of basic contract law."[216]

Supporters of web scraping should not be tempted to read the *Bright Data* case to support the sweeping proposition that all scraping of publicly-viewable data does not violate the platform's terms of service. Rather:

> the case serves to reaffirm the broad general ability to webscrape publicly available portions of websites where an account login/password has not been utilized. Where an account login/password may have been utilized in unrelated contexts, the precise contractual terms agreed to by the user will be important to determine whether webscraping is allowable. While current terms are unlikely to expressly restrict webscraping while logged out (or after termination of an account), website owners may react to the court's decision in Meta v. Bright Data by attempting to modify their terms to broadly restrict these activities through specific language to this effect. It remains to be seen whether such broad restrictions on webscraping activity (while logged out) would be upheld by courts even if terms of use are drafted to maximum effect.[217]

[212] Ahmed Eissa, Daniel Levin, Stacy Bradenburg, and Sheri Pan, ZwillGenBlog, hiQ v. LinkedIn: Breach of Contract and CFAA Claims Proceed to Trial, (last visited December 8, 2023), https://www.zwillgen.com/alternative-data/hiq-linkedin-breach-contract-cfaa-trial/.

[213] Id.

[214] Meta Platforms Inc. v. Bright Data Ltd., Case No. 3:23-cv-00077-EMC (N.D. Cal.).

[215] Order, Meta Platforms Inc. v. Bright Data Ltd. (January 23, 2024), https://www.courthousenews.com/wp-content/uploads/2024/01/meta-platforms-v-bright-data-ruling-motion-for-summary-judgment.pdf.

[216] Id.

[217] Scott H. Moss, Boris Liberman, George Danenhauer, and Michael J. Scales, Lowenstein Sandler, Meta v. Bright Data Ruling Has Important Implications for Webscraping Activities by Investment Advisers (February 15, 2024), https://www.lowenstein.com/news-insights/publications/client-alerts/meta-v-bright-data-ruling-has-important-implications-for-webscraping-activities-by-investment-advisers-im).

5. Misappropriation

The Northern District of California denied hiQ's motion to dismiss LinkedIn's misappropriation claim.[218] The elements that California required to state a claim for misappropriation analysis are as follows: "(1) the plaintiff has invested substantial time and money in development of its … 'property'; (2) the defendant has appropriated the [property] at little or no cost; and (3) the plaintiff has been injured by the defendant's conduct."[219] In this analysis, a court will determine whether defendant is "free-riding" and taking advantage of plaintiff's intellectual property without consent.

Although the Ninth Circuit never officially ruled on the misappropriation question in the hiQ case, qualifying the CFAA as an "anti-intrusion statute and not as a misappropriation statute," it stated in dicta that this cause of action may be available to plaintiffs whose data has been scraped without permission.[220]

6. Unjust Enrichment and Conversion

Similarly, the Ninth Circuit did not analyze LinkedIn's unjust enrichment or conversion claims, but stated in dicta that they may be available for web scraping.[221] Under California law, the elements of a claim of unjust enrichment are defendant's receipt of a benefit and unjust retention of that benefit at the plaintiff's expense,"[222] and the elements of a conversion claim are plaintiff's ownership or right to possession of the property at the time of the conversion; defendant's conversion by a wrongful act or disposition of property rights; and damages to plaintiff.[223]

A 2023 class action lawsuit, *J.L. v. Alphabet,* which is still pending, involves (among many other claims) a claim of conversion with respect to scraped data to train AI models.[224] Plaintiffs state that they have a property right in their personal data and information, which they generated, created, and/or provided through various online platforms. In light of this purported ownership, plaintiffs claim that they have the right to "possess, use, profit from, sell, and exclude others from accessing or exploiting that information without consent or remuneration."[225]

---

[218] See HiQ Labs, Inc. v. LinkedIn Corp., 273 F.Supp.3d 1099, 1104 (N.D. Cal. 2017).

[219] Balboa Ins. Co. v. Trans Global Equities, 218 Cal.App.3d 1327, 1334 (1990).

[220] See hiQ, 938 F.3d at 1016.

[221] See hiQ, 938 F.3d 985.

[222] MH Pillars Ltd. v. Realini, 277 F. Supp. 3d 1077 , 1094 (N.D. Cal. 2017).

[223] See, e.g., Firoozye v. EarthLink Network, 153 F. Supp. 2d 1115 (N.D. Cal. 2001); Burlesci v. Petersen, 68 Cal.App.4th 1062 (App. 1st Dist. 1998); Oakdale Village Group v. Fong, 43 Cal.App.4th 539 (App. 3d Dist. 1996); Lee v. Hanley, 61 Cal.4th 1225 (2015).

[224] See J. L. v. Alphabet Inc. (3:23-cv-03440).

[225] See, e.g., Calhoun v. Google, LLC, 526 F. Supp. 3d 605, 635 (N.D. Cal. 2021) (recognizing property interest in personal information and rejecting Google's argument that "the personal information that Google allegedly stole is not property"); In re Experian Data Breach Litigation, SACV 15-1592 AG (DFMx), 2016 U.S. Dist. LEXIS 184500, at *14 (C.D. Cal. Dec. 29, 2016) (loss of value of personal identifying information is a viable damages theory); In re Marriott Int'l Inc. Customer Data Sec. Breach Litig., 440 F. Supp. 3d 447, 460-61 (D. Md. 2020) (noting "[t]he growing trend . . . to recognize the lost property value of this [personal] information."); Simona Opris v. Sincera, No. 21-3072, 2022 U.S. Dist. LEXIS 94192, at *20 (E.D. Pa. May 23, 2022) (collecting cases). See also Ajemian v. Yahoo! Inc., 84 N.E. 3d 766 (Mass. 2017) (an email account is a "form of property often referred to as a 'digital asset.'"); Eysoldt v. ProScan Imaging, 957 N. E. 2d 780 (Ohio App. 2011) (permitting action for conversion of web account as intangible property).

Plaintiffs allege that defendants "unlawfully collected, used, and exercised dominion and control over" their personal and private information (e.g., photos from dating websites, Spotify playlists, and books) for use in training defendants' AI model without permission or authorization.[226]

7. Breach of Privacy

Again, without analyzing the claim, the Ninth Circuit in hiQ stated that "breach of privacy [claims], may also lie."[227] In the wake of hiQ, several class action lawsuits have been filed alleging privacy violations associated with generative AI tools,[228] including *J.L. v. Alphabet, Inc.*, *PM v. OpenAI,* and *Brantley v. Prisma Labs, Inc.*[229] In each of these cases, plaintiffs allege the AI tools were illegally trained on plaintiffs' data that was web-scraped from the internet.[230] In the PM case, OpenAI's web scraping practices are alleged to violate users' privacy rights under the Electronic Communications Privacy Act, the California Invasion of Privacy Act, and the Illinois Biometric Information Privacy Act ("BIPA").[231] These violation are alleged to have occurred when OpenAI used biometric identifiers (e.g., scanned facial geometry of faces found in photographs) to train algorithms to power its DALL-E AI-infused image generation tool without informing the individuals in writing of the purpose for the collection of this data.[232] In Brantley, plaintiffs argue that Prisma Labs, Inc.'s mobile app violated their privacy rights by scraping their photographs from the internet without their consent or notice.[233]

As cases like the aforementioned continue to percolate throughout the courts, breach of privacy actions are ripe for plaintiffs to utilize as tools against artificial intelligence models and their web scraping practices.

VI. Conclusion

Artificial intelligence, particularly with respect to technical and operational matters, has had a profound impact on digital advertising for a number of years. The relatively recent proliferation of generative-AI is already having a material impact on the creation of content for both publishers and brands and is set to transform the industry – and many other industries – for years to come.

[226] See J. L. v. Alphabet Inc. (3:23-cv-03440).

[227] See hiQ, 938 F.3d 985.

[228] See Michael Benedetti, BloombergBNA, ANALYSIS: Early Generative AI Lawsuits Pinpoint Privacy Concerns, (last visited December 12, 2023), https://www.bloomberglaw.com/bloomberglawnews/bloomberg-law-analysis/X3E0752K000000?bna_news_filter=bloomberg-law-analysis#jcite.

[229] Id.

[230] Id.

[231] Id.

[232] Id.

[233] Supra note 184.

Described by one expert as "the coming wave,"[234] it is predicted that AI's "transformative applications will both empower humankind and present unprecedented risks."[235] As this whitepaper has shown, there are numerous benefits and risks associated with generative-AI. Publishers and advertisers are well-advised to create, monitor, and enforce acceptable use policies governing their use of artificial intelligence to ensure that: (i) the quality and accuracy of their content and advertisements – including those created by their personnel or in-licensed from third parties – meet or exceed the level of human-generated content; (ii) journalistic standards and brand safety guidelines are followed; and (iii) they are being transparent in their use of AI, complying with their contractual obligations and the panoply of laws, rules, and regulations being promulgated every day, and reducing – to the maximum extent possible – bias.

The transformative power of AI has also inspired an international race to address the legal and ethical concerns inherent in AI. At the time of publication, over two dozen jurisdictions have issued AI regulations, with the most comprehensive regulation to date issued by the EU. Similar to the GDPR, the EU's Artificial Intelligence Act ("AI Act") applies to businesses that provide or develop AI tools that are marketed or used in the EU.[236] Importantly, the AI Act divides AI systems into a four-tier, risk-based system: Unacceptable Risk, High Risk, Limited Risk, and Minimal Risk. AI systems presenting an unacceptable risk include those that can be used to exploit or discriminate against an individual based on characteristics such as their age, race, or gender. These systems are prohibited outright. High risk systems span a wide array of topics, including those used for biometric identification systems, law enforcement, or education. These are tightly regulated and must adhere to, amongst other things, principles of transparency and human oversight. Similarly, Limited Risk refers to AI systems that present a risk of manipulation in their interactions with humans, as is the case with chatbots, and thus require a high degree of transparency that fosters user awareness. Finally, Minimal Risk AI systems refers to those applications that are already in wide use, such as spam filters, and which present little to no risk to users. Generative AI does not easily fit into the EU AI Act's risk schema, and it's important to assess generative AI practices based on a case by case basis. On its face, the EU's risk-based approach seeks to strike a delicate balance between stringent regulation and fostering technological innovation, though the effects remain to be seen.

The AI genie is out of the bottle and there is no going back. The key for advertisers, agencies, publishers, and other stakeholders in the digital advertising ecosystem is to utilize the technology in a responsible, transparent, and compliant manner.

[234] Mustafa Suleyman and Michael Bhaskar, The Coming Wave: Technology, Power, and the 21st Century's Greatest Dilemma, Crown Publishing Group (2023).

[235] Id. at p. vii.

[236] Texts Adopted (EU) 2024/0138 European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))