

**THE CONFLUENCE OF LAWS INVOLVED IN THE DEVELOPMENT AND USE  
OF GENERATIVE ARTIFICIAL INTELLIGENCE**  
and  
A BALANCED, ONE SIZE DOES NOT FIT ALL, APPROACH TO SOME RELATED  
INTELLECTUAL PROPERTY ISSUES

Written by  
Matthew Neco and Jason Peterson<sup>1</sup>  
for GoDigital Media Group, LLC<sup>1</sup>

## Need to Know Items

This paper includes:

- A simple model for GAI and copyright regarding (i) training, (ii) prompting, and (iii) output.
- A simple model for property rights in the context of GAI is: Control, Compensation, and Credit.

We recommend:

- That property rights should include copyright and personal digital replica rights.
- Statutory licensing for training and direct licensing for outputs.
- Standardized training AI.txt machine readable file that can be used by rights-holders and online service provider that sets forth training licensing terms or conveys a message that the rights-holder is amenable to negotiating licensing terms.

---

<sup>1</sup> © 2023 GoDigital Media Group, LLC. **Most of this work where we hold copyright licensed pursuant to [Creative Commons Attribution 4.0 International License](#).** Cite as “Neco, Matthew and Peterson, Jason, for GoDigital Media Group, LLC “The Confluence of Laws Involved in the Development and Use of Generative Artificial Intelligence and a Balanced, One Size Does Not Fit All, Approach to Some Related Intellectual Property Issues.” Portions not licensed by us are quotes, abstracts, etc., not licensed by existing third party rights-holders, or where third party licenses are not under [Creative Commons Attribution 4.0 International License](#), though we believe that our uses are fair uses and not infringements.

This work-in-progress paper outlines some of the evolving thinking of GDMG\* regarding GAI tools. Except when quoting a third party, the use of “we” or “our” is intended to denote GDMG,\* not necessarily the human authors. Some of what is discussed herein is intended to provoke thinking and discussion. GDMG,\* and the human authors of this paper, reserve their rights to make corrections and modifications, to disavow and disclaim recommendations, policies and positions, and to pivot, among other things, including because of changing or evolving thinking, circumstances and considerations (including, for example, because of proposed or newly enacted laws, or regulations, and executive orders, as well as negotiated agreements, lawsuits or regulatory actions and outcomes premised on any of the foregoing). GDMG and one or more of its business units are associated with a number of collection societies and other industry groups, however, GDMG,\* and the human authors, do not and may not necessarily endorse or agree with some or any of the positions of them. Nothing in this paper should be used as evidence or cited against GDMG\* or any of its human authors, investors, executives, employees, etc., including in any disputes, discovery, inquiries, etc.

\*and/or one or more of its business units

Jason can be contacted by emailing [jason@godigitalmg.com](mailto:jason@godigitalmg.com)

- Open source blockchain based solution to provide copyright management information to AI companies for training and licensing and from which to provide provenance information to enable downstream business models and compliance with existing and future laws and regulations, including Federal Trade Commission truth in advertising laws.
- That the Copyright Office design standards for copyrightability in the context of prompt engineering.
- A standardized API between GAI companies and the Copyright Office to provide the necessary information to determine if prompt engineering gives rise to copyrightability.
- The Artist Centric approach to licensing digital service providers as pioneered by Universal Music Group in its deal with Deezer as a first step in a licensing model enabling GAI music and human authored music to live side by side in the consumer experience.
- That performers' voices and likenesses must only be used in outputs with their consent and fair market compensation for specific uses.
- That every human's name, image, voice, likeness, and individual style (in some circumstances) be a property right, perhaps as suggested in a better conceived version of the discussion draft of the NO FAKES act.
- That Section 230 of the Communications Decency Act be amended to clarify that state right of publicity and privacy-based appropriation claims are not immunized by §230 unless and until there is new Federal pervasive legislation that preempts state law.

## Background

GoDigital Media Group, LLC<sup>2</sup> (GDMG) embraces and uses many technological advances provided by traditional artificial intelligence (AI), as well as generative AI (GAI). But, like many, we have some concerns about GAI. As a somewhat new-to-the masses technology, GAI calls upon a wide variety of laws and regulations (the “confluence of laws”) across a multitude of jurisdictions that include state, federal and international forums – and these different laws and regulations often are not in complete harmony (the “conflict of laws”).

This paper touches upon some of the primary issues surrounding the confluence of laws in GAI, Foundation Models, and Large Language Models (LLMs). This includes the training of GAI tools through the collection (scraping), ingestion, and use of content and data, including

---

<sup>2</sup> GoDigital Media Group, LLC is a privately held company. Its subsidiaries are an ecosystem connecting content, community, and commerce at consumer passion points. GoDigital's purpose is to inspire happiness. GDMG's interests include Cinq Music, Sound Royalties, NGLmitu, Eastern Mountain Sports, Bob's Stores, and YogaWorks. We use GDMG as shorthand for GoDigital Media Group, LLC, individually and the entities that GoDigital Media Group, LLC has a business interest in.

works of human authorship and in-copyright works. We also examine the outputs of using GAI tools. We look at these issues primarily in connection with some of the laws, regulations, and norms around copyright, and name, image, likeness, voice & style (NILVS). We provide an overview of the legal and regulatory schemes and norms of the United States of America (US), as well as an even briefer overview of those of the European Union (EU) and its member states, and some other countries that are not members of the EU.<sup>3</sup>

In addition to looking at the framework of some laws and regulations that apply, we examine the implications for the marketplaces of human authors and the creative copyright and content industries. We suggest some possible solutions to some of the issues raised in this paper.

Some of the issues and objectives discussed are at tension with each other, as are some of the legal and regulatory schemes of the US and different international jurisdictions. As an example of such tension, humanity has an interest in GAI tools producing (or assisting in producing) outputs that are transparent, accurate, informative, educational, entertaining, unbiased, and non-discriminatory, among other things. Though a primary approach of this paper is the focus on entertainment, infotainment or edutainment, we must not and do not forget that there are other high stakes issues in play, such as in relation to medicine. To achieve anything close to these goals it appears that it is in the best interests of society for Data Hosts and GAI Model Creators to have the largest possible universe of data and content available for ingestion and then the training of GAI tools. This is apparently in tension with some of the goals of some very large populations of copyright holders, collection societies, and trade and lobbying groups, many of which traditionally have taken, and currently in connection with GAI, take the position that it is rightsholders,<sup>4</sup> and rightsholders alone, who

---

<sup>3</sup> Although LLMs, Foundation Models, and GAI models and tools have been in development and use for many years, they only recently entered the zeitgeist of the general population. Many of the issues touched upon in this paper have been and are being addressed at length by researchers, scholars and academics, practitioners, rights-holders, think tanks, advocacy groups, public-interest groups, regulators, and policy makers, at great length, some over many years, including in lengthy and complex scholarly peer reviewed academic papers and at conferences, legislative and regulatory “listening sessions” and hearings, discussing, debating, and disagreeing about applicable existing and desirable laws, norms, and computer science, among other things.

Although we may be a bit late to the game in circulating or publishing a draft of this paper, we are playing this infinite game. We have ingested and learned from and continue to learn from a relatively small portion of the corpus on just the few issues touched upon in this paper. We suspect that might be true of others who or that have not devoted a substantial amount of time over the last number of years to these issues. Notwithstanding that, we have engaged in some research, learning, and thinking, and we are dipping our toes in the waters. Although this is just the beginning of our own journeys, we will continue to study, learn, discuss, debate, and disagree about these issues, and may update this paper over time, or write additional papers on some of these issues, even as we consider how we may use or develop, and profit from, GAI tools in our ecosystem. Further, we are not experts in many of the fields discussed in this paper, and this paper has not been peer reviewed, so we might occasionally get something incorrect, for example in the fields of how GAI systems are trained or function, or rights of publicity. Also, some portions are not yet fully developed. We welcome corrections and constructive criticism. Finally, we recognize that this paper could be updated from this 11/28/23 draft as a lot has transpired in the GAI world since then.

<sup>4</sup> Here, mostly the copyright holders or owners, or exclusive licensees of certain or all rights in a work, more so than collectives or collection societies and non-exclusive licensees with no or very few rights to control what may be done with works. Also, “authors,” and “rights-holders” are not necessarily synonymous.

should have exclusive control over what may be done with in-copyright works of authorship<sup>5</sup> – how, where, when, why, by whom, and for how long, content may be used.<sup>6</sup>

We attempt to take a somewhat medium to longer term perspective on the prevailing goals and benefits of encouraging human creativity. These views favor iterative approaches allowing for research, technological experiments and improvements, and well-reasoned consideration for positive and negative effects of proposed changes and amendments to laws and regulations over, “let’s get ahead of the ‘problems,’ without another moment’s delay and enact sweeping new laws, regulations, executive orders,<sup>7</sup> and directives.”<sup>8</sup> Laws and regulations very rarely keep up with technological innovations. We recognize, however, that although there may be existing laws and regulations subject to interpretation by Courts and regulators, these laws and regulations may not adequately address certain issues that have arisen and are rising with regard to GAI. There is some legitimate urgency in addressing some of these issues *post haste*. Some of these issues are well outside the scope of this paper, but, for example: there are legitimate concerns that deep fakes will, among other things, contain misinformation that might influence outcomes of democratic elections, and we need to address this now and not wait until after elections – though we must be vigilant about protecting free speech rights, among other things.

## Executive Summary

We suggest some solutions to the following issues:

- Issue 1: **Copyright** - What the rights of copyright holders are, might be, or should be, including:
  - whether there is or may be **copyright infringement**
    - by the unlicensed **scraping** and use of copyrighted works in LLMs and Foundation Models, and the training of GAI tools.

---

<sup>5</sup> We may shorten the phrases “in-copyright works of authorship,” and “works of authorship,” to “works” or “content.”

<sup>6</sup> Excepting, to a certain extent, those rights-holders who license their works under Free and Open Source Software licenses, and certain Creative Commons types of licenses.

<sup>7</sup> For example, on October 30, 2023, U.S. President Joe Biden issued an Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. See, the “Fact Sheet” at <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/> and the EO at <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>, as well as the new AI.gov website.

<sup>8</sup> There have been several hearings before various subcommittees of the United States Senate and House of Representatives regarding GAI and intellectual property and deepfakes and whether there should be a federal right of publicity law. The United States Copyright Office has had several listening sessions and webinars on GAI and copyright, and has published a Notice of Inquiry and Request for Comments in connection with GAI, and has included in that NOI questions adjacent to copyright, specifically about deep fakes and whether there should be a federal right of publicity law. The USCO has a webpage devoted to GAI. See, [Copyright and Artificial Intelligence | U.S. Copyright Office \(https://copyright.gov/ai/\)](https://copyright.gov/ai/).

- by the unlicensed presence of content in **outputs** that might include elements of pre-existing copyrighted works, whether in fact, or with substantial similarity.
    - whether the defenses or doctrines of what is known in the United States as **fair use**, or in some other countries as **fair dealing, or text and data mining limitations & exceptions** apply in considering the foregoing and, therefore, licenses are not required.
  - regardless of whether **fair use, fair dealing, or text and data mining exceptions** apply, should there be **licensing** schemes in connection with ingestion and output, and whether licensing should be on a **negotiated or blanket compulsory** basis.
  - whether and to what extent copyright holders have or should have the right to:
    - **consent to the ingestion** of content by LLM and GAI developers, and if there is or should be that right whether it should be exercised by **opt-in or opt-out**, and how.
    - in addition to copyright holders, whether website operators might use **contracts** (terms or use / terms of service) to attempt to **prohibit** scraping.
    - delineate the circumstances under which their works may and may not be used in ingestion, and the creation of outputs.<sup>9</sup> For example, might a copyright holder have rights to prohibit the use of their work in output they find offensive, in conflict with their values, or illegal.
    - receive **credit** or attribution, including within copyright management information.<sup>10</sup>
- Issue 2: **Right to Identity** - What the rights of humans might or should be involving **identity** (such as name, image, likeness, voice, and style) and **publicity and privacy**, as well as **defamation**, and **deep fakes**.
  - **Thoughts:** On October 12, 2023, several U.S. Senators released a bi-partisan discussion draft of their Nurture Originals, Foster Art, and Keep Entertainment Safe (NO FAKES) Act intended to protect the voice and visual likenesses of individuals (and entities that have licenses) from unauthorized uses through GAI. We believe many of the proposals are good, including that protection does not apply to works of parody or satire, news pieces or documentaries, and other similar categories.
- Issue 3: **Bias** – A primary issue is mitigating against biases and discrimination that might be present in the output or results of using GAI tools.

---

<sup>9</sup> For these purposes we will assume a copy or facsimile of some portion of a copyright holder’s work is actually factually present in output, or some portion of the output appears to contain some portion of expression that may be perceived, reproduced, or communicated, which is substantially similar to some portion of a copyright holder’s work. More on these assumptions within this paper.

<sup>10</sup> We believe that something akin to Section 1202 of the Copyright Act (Pertaining to copyright management information) might apply. See, [Chapter 12 - Circular 92 | U.S. Copyright Office](#)

- **Thoughts:** This can only be done if there is a huge corpus of material ingested, de-duplicated, and reviewed, which requires prohibiting rights-holders from opting-out of ingestion, or “poisoning.”<sup>11</sup> Content to be ingested must not contain information subject to privacy laws and regulations, and the content must not be illegal, false, fake, biased, etc. We have however heard unconfirmed information from reliable sources that certain AI and GAI training sets have appeared to incorporate intentionally biased material – a new form of malware. In an era where the skillset of the future is being able to ask the right questions instead of rote memorization of facts, bias in the answers and outputs we receive from GAI has tremendous implications for society because users may not have context to measure the GAI’s answers against.
- Issue 4: **Work Product Ownership** - What the rights of those who perceive or receive outputs might or should be:
  - **Transparency.** That is, the right to be informed that:
    - the work was created by, or the creation was assisted by, a GAI tool, and how granular such provenance information might be.
    - whose copyrighted works are or might be included in the outputs, and how granular such provenance information might be, including as pertains to copyright management information.
    - when output is or contains deep fake representations, and how granular such information might be.
    - when output is or contains the representations or depictions of a human, whether such representations are of a real human (living or dead), or a simulated human, and how granular such information might be.
  - **Copyrightability**
    - whether and to what extent **prompts** (including text, illustrations, audio, audio video, formulas, software code, etc.) created or authored and used with GAI tools might or should be **copyrightable** if (a) the prompts are human created, or if (b) some or all of the prompts are generated by the GAI tool.
    - whether and to what extent **outputs** generated by or with the assistance of GAI tools are or should be **copyrightable**, and to what extent human contribution of creative work is adequate or necessary in order for some portion or all of the GAI assisted or created output to be copyrightable.

## Definitions

The following are brief, general descriptions:

**Traditional or Classical AI:** Traditional or Classical Artificial Intelligence, also known as rule-based AI, has been described as a form of artificial intelligence that operates based on

---

<sup>11</sup> See, e.g., [This new data poisoning tool lets artists fight back against generative AI | MIT Technology Review](#).

predefined rules or algorithms. Think of it as an intricate flowchart where the AI navigates through a specific path determined by the input it receives. A common example of this in everyday life is a customer service chatbot. These chatbots have a predetermined list of responses and select what may be the most fitting one based on the keywords identified in your message. Let's consider some instances of AI that are not generative, such as facial recognition on mobile devices. This type of AI operates by analyzing the unique features of your face and comparing them to a stored image. It's important to note that this process does not involve the creation of new content. Instead, it strictly adheres to a predetermined set of rules to recognize your face.

Traditional AI solves specific problems or performs specific tasks and focuses on pattern recognition, data classification and decision making. In music, for example it is traditional AI that has been used by Digital Service Providers, such as YouTube, to power recommendation algorithms, it is used to improve sound quality in production and mastering, and it is used by some Collective Management Organizations to help identify copyright owners.

**Generative AI:** Generative AI is a more recent form of AI that has the capability to produce new content. Unlike other forms of AI that strictly adhere to a set of rules, generative AI is capable of 'learning' and subsequently generating something new from what it has 'learned'. For instance, generative AI can be seen in action when an AI 'composes' a poem or 'creates' an illustration. The distinguishing feature of generative AI is its ability to generate novel (or some would say, derivative) creations, rather than merely following a set of pre-established instructions.

**The Actors:** We find the **actors** defined by Henderson, et al., in the paper "Foundation Models and Fair Use" to be helpful. We've modified the definitions a bit, so don't blame the authors of that paper if we've gotten anything incorrect.<sup>12</sup>

- The **data creator** (*aka author*) creates data (works of authorship / content) that a GAI model might be trained on. (There might be a copyright-holder involved who might or might not be the creator of the work.)
- The **data curator** collects data (by various means, including scraping, or being provided with the data), usually created by many data creators (authors). These include LAION ([FAQ | LAION](#)) and [Common Crawl](#).
- The **data host** distributes or makes available data (the works) that a GAI model is trained on. The data host might be a data curator or not.
- The **model creator** "**trains**" the GAI model on this data.
- The **GAI model deployer** hosts a GAI model and provides access to it via an API or otherwise, potentially creating revenue from serving the GAI model. An example of a GAI model deployer might be the user accessible version of ChatGPT, or using the "chat" feature in MSFT's Bing / Edge Browser, or "Bard" via Google, or Firefly by Adobe.

---

<sup>12</sup> Henderson, Peter and Li, Xuechen and Jurafsky, Dan and Hashimoto, Tatsunori and Lemley, Mark A. and Liang, Percy, Foundation Models and Fair Use (March 27, 2023). Stanford Law and Economics Olin Working Paper No. 584, Available at SSRN: <https://ssrn.com/abstract=4404340>, <http://dx.doi.org/10.2139/ssrn.4404340>, [delivery.php \(ssrn.com\)](https://creativecommons.org/licenses/by/4.0/) ©2023 Henderson, Li, Jurafsky, Hashimoto, Lemley, & Liang. License: CC-BY 4.0, see <https://creativecommons.org/licenses/by/4.0/>. See, also, the shorter version of the paper available here: [Policy Brief | Foundation Models and Copyright Questions | Stanford HAI](#) (<https://hai.stanford.edu/policy-brief-foundation-models-and-copyright-questions>) linking to [Foundation-Models-Copyright.pdf \(stanford.edu\)](#) <https://hai.stanford.edu/sites/default/files/2023-11/Foundation-Models-Copyright.pdf>

- The **GAI model user** uses the GAI model for downstream tasks creating output (the **GAI output**), potentially creating revenue with the output of the GAI model.
- The **GAI output user** distributes, displays, exhibits, publishes, performs, etc., the GAI output.

## A Brief Description of the Process (as we understand it)

The **Data Curator**, in the process of collecting or scraping "data," actually **downloads and saves the data** (code, text, images, audio, video, etc., often in-copyright works of authorship), to a hard drive, whether that drive is local or "cloud" based, such as on AWS. There has been a **reproduction** of the data by the Data Curator.

In some instances, the Data Curator might have attempted in good faith to only capture or receive data / works that are in the public domain (not in-copyright) or are purportedly licensed per a permissive Creative Commons type license or FOSS license. Or in a situation such as with Adobe's Firefly, actually licensed to Adobe by the Data Creator (author / photographer), or where a platform might have an explicit or implicit / interpretable license from the poster / uploader (the Data Creator) to use the works to train the GAI model, whether the user knows or knew this or not. (Meta claims no Meta user data was used in Lama2.<sup>13</sup>)

The Data Curator might be the **Data Host** or another party / entity. If another entity the Data Curator makes some or all of the copies of data available to the Data Host, and if the Data Host is a party that is different from the Data Curator the Data Host **copies and saves the data** / works to a hard drive, not merely as transitory RAM or cache copies. These copies are actual **reproductions** of the saved data, or works of authorship. In some instances the Data Host might attempt to filter out or otherwise segregate data / works that are not in the public domain or not permissively licensed before the next step in the process. Keeping copies of the data / works may be good, for deduplication, and to mitigate against bias, discrimination, etc.<sup>14</sup>

The data host distributes or makes available to the **GAI model creator** data that a GAI model is **trained on**. Many computer scientists assert that the GAI model creator does not actually copy, **reproduce**, or download, even for a fraction of time, any of the data / works. For example, it is asserted the GAI model creator does not copy, reproduce, download a song written and performed by Taylor Swift or a poem by Maya Angelou. We disagree on the premise a computer cannot detect the presence of information without at least ephemerally caching it.

---

<sup>13</sup> See, p. 9, Llama 2: Open Foundation and Fine-Tuned Chat Models, available at [10000000\\_662098952474184\\_2584067087619170692\\_n.pdf \(fbcdn.net\)](https://arxiv.org/pdf/2307.09288v2.pdf).

<sup>14</sup> **LAION** has written that it discards the copies of the photos: "LAION datasets are simply indexes to the internet, i.e. lists of URLs to the original images together with the ALT texts found linked to those images. While we downloaded and calculated CLIP embeddings of the pictures to compute similarity scores between pictures and texts, we subsequently discarded all the photos. Any researcher using the datasets must reconstruct the images data by downloading the subset they are interested in. For this purpose, we suggest the img2dataset tool."

The **GAI model** that the **GAI model user** uses does not contain, or link to, or call to, copies of even snippets or pixels of the actual original data / works (e.g., a sentence or a phrase from a Taylor Swift song, such as "Players gonna play, Haters gonna hate," or a line from a Maya Angelou poem, such as "the free bird thinks of another breeze").

After the **GAI Model Creator** has **trained** the GAI model, GAI Model Creator deploys or makes the GAI model available to the model user via a user interface or an API.

In some instances a model user might use the GAI model or tool to generate (or assist in generating) something that might regurgitate verbatim (or almost verbatim), or resemble (be substantially similar to) some or all of a distinct work of authorship, such as an entire section of a book, or appear to be Drake and the Weeknd performing a song that might be substantially similar to a song that Drake and/or the Weeknd composed or would compose and performed as if they performed it. In other instances, it might be the case that no portion of the output would be even remotely similar to the works of any authors, or at least not perceived so by the average or ordinary person. In some instances the output might be like a phonorecord that contains many samples, some of which might not be recognized by ordinary humans (e.g., [Beastie Boys - B-Boy Bouillabaisse \(1989\) https://youtu.be/EwGjVsCSOsU](https://youtu.be/EwGjVsCSOsU) or [Unkle Intro \(Optional\) Samples \(with subs\) https://youtu.be/oRwnUM10mf4](https://youtu.be/oRwnUM10mf4), or a "loop" that has been sampled extensively, such as the 6 second [Amen Break](#)).

## Confluence of Laws

### 1. COPYRIGHT ISSUES WITH REGARD TO THE INPUTS USED IN LLM & GAI TOOL DEVELOPMENT:

#### **Infringement of Exclusive Copyrights, Fair Use Under US Law, and Text & Data Mining Exceptions in The EU and Select Countries**

Copyright holders in the United States and broadly around most of the world under the Berne and Rome Conventions generally have what amounts to the exclusive right to reproduce, distribute, publicly perform, publicly display and make works derivative of their owned or controlled works. In order for GAI models to study and learn from content they or their precursor actors/programs must at least make ephemeral time limited reproductions of the content in the CPU cache and RAM of their servers.<sup>15</sup> It is the general belief that popular GAI

---

<sup>15</sup> There are some who opine that works are not reproduced or copied to train GAI tools. For example, at a hearing on July 12, 2023, before the U.S. Senate Judiciary Committee's Subcommittee on Intellectual Property, which has very informative written testimony, oral testimony, and statements and questions by Senators. <https://www.judiciary.senate.gov/artificial-intelligence-and-intellectual-property-part-ii-copyright>. The written submission and oral testimony by Law Professor Matthew Sag, states, in part "One of the most common misconceptions about Generative AI is the notion that training data is routinely "copied into" the [GAI] model. Machine learning models are influenced by the data, they would be useless without it, but they typically don't copy the data in any literal sense. In very rare cases when they do copy the training data—something computer scientists call "memorization"—that is regarded as a bug to be fixed, not a desirable feature. [https://www.judiciary.senate.gov/imo/media/doc/2023-07-12\\_pm\\_-\\_testimony\\_-\\_sag.pdf](https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_sag.pdf) .

We believe there is a multi-step process. First the Data Curator scrapes and ingests works. Then the GAI Model Creator converts the works into some mathematical formulae. Then it trains the GAI tools or models the formulae but not any actual parts of works. GAI tools are not fed whole sections of books, for example, or even sentences. The GAI tool makes predictions based upon the layers of formulae it has been trained on. These are

tools and models are generally trained on very large datasets of media (such as publicly available images, texts, audio-video, and sound recordings) scraped or copied from the Internet and digital service providers such as YouTube. Much of this content is in-copyright, whether registered or not, with the balance in the public domain (not in-copyright).<sup>16</sup> Because assembling these training datasets most likely involves making reproductions and distributions of copyrighted works, this has raised the question of whether first the scraping and ingestion, and second the process of making mathematical representations of the copyrighted works for training GAI tools, infringes any of the copyright holders' exclusive rights.

This reproduction and distribution of data over the network that represents copyrighted works used to train GAI tools may or may not be allowable as **Fair Use** under the Copyright Act in the US,<sup>17</sup> or in the EU pursuant to what is known as the **Text & Data Mining** (TDM)

---

two instructive explainers: How AI chatbots like ChatGPT or Bard work – visual explainer, <https://www.theguardian.com/technology/ng-interactive/2023/nov/01/how-ai-chatbots-like-chatgpt-or-bard-work-visual-explainer>  
 Generative AI Exists Because of the Transformer, <https://ig.ft.com/generative-ai/>

<sup>16</sup> “**Public Domain**” is a legal term that basically means something is not protected by copyright, such as, because the copyright term has expired and the work is no longer “in-copyright,” or in some instances the work was deemed “authored” by the United States Government. “Public Domain” does **not** mean something is not proprietary and not protected by copyright just because it is **publicly available** via the internet.

For example, just because a book is in the public library and publicly available that does not mean it is in the public domain and not proprietary and not protected by copyright laws. An example of what is in the public domain is a musical composition written by Beethoven, because Beethoven authored it over 100 years ago, however, a recent sound recording – a recent recording by a symphony orchestra, for example - of a Beethoven musical composition is probably not in the public domain because the recent sound recording of the performance itself is likely protected by copyright. Like many matters regarding copyright, what is or is not in the public domain may be difficult to ascertain. The rules are complex and vary from jurisdiction to jurisdiction. Many works authored before 1928 may now be in the public domain because copyright protection has expired, at least in the US. (See, e.g., Duke University School of Law’s [Center for the Study of the Public Domain](#), and these Wikipedia pages discussing the [public domain](#) and [list of countries’ copyright lengths](#).)

Further, some GAI tool developers claim that much of the data it used is from [Wikipedia](#) and the entries there are available under the [Creative Commons Attribution-ShareAlike License](#). Creative Commons also has a license type that allows authors to dedicate their works to the [public domain](#). Creative Commons does not warrant that just because a work or purports to be dedicated to the public domain that the dedication was made by the actual rights holder, or that an entry or portion of an entry was not an infringement. Similarly, most Free and Open Source Software licenses disclaim assurances that portions of the code do not infringe upon the rights of any third party.

<sup>17</sup> As most readers know, Fair Use is legal doctrine in the US intended to promote freedom of expression by permitting the unlicensed use of copyright-protected works, in whole or in part, in certain circumstances. Determining whether uses are fair uses may be very fact specific and the ultimate determination is often made in lawsuits. For example, just because there is news about someone famous it would probably not be fair use to copy a photograph of that person you find on Instagram by cutting or copying that photo and pasting that photo in a news story that you publish without the permission of the rights holder of the photo. Further, a recent United States Supreme Court majority opinion in what is colloquially known as the Warhol v. Goldsmith case decided May 18, 2023, addressed a very narrow aspect of fair use, and some of the Justices disagreed with others about whether the uses in question in that case were or were not fair uses under the law. The Supreme Court’s opinions are here: [https://www.supremecourt.gov/opinions/22pdf/21-869\\_87ad.pdf](https://www.supremecourt.gov/opinions/22pdf/21-869_87ad.pdf).

The US Copyright Office has published this, about Fair Use:

**Exceptions** and Limitations of Articles 3 and 4 of the Directive on Copyright in the Digital Single Market,<sup>18</sup> and the laws of other select countries, such as the United Kingdom. The laws of many countries allow for TDM, to some extent, for research, though not necessarily for commercial uses and exploitation. The laws of some countries also allow for text and data mining for commercial purposes. See the footnotes for an open access law review article that shows countries that might have TDM and TDM-type exceptions and limitations.<sup>19</sup>

In addition, with regard to training, besides the Fair Use doctrine in the US, and TDM exceptions where they apply, there might be allowances and exceptions for training under what may generally be referred to as “**Fair Dealing**” allowances and exceptions in some 40 countries in the world. **Fair dealing** is a limitation and exception to the exclusive rights granted by copyright law to the author of a creative work. Fair dealing is found in many common law jurisdictions.

---

[Section 107 of the Copyright Act](#) provides the statutory framework for determining whether something is a fair use and identifies certain types of uses—such as criticism, comment, news reporting, teaching, scholarship, and research—as examples of activities that may qualify as fair use. Section 107 calls for consideration of the following four factors in evaluating a question of fair use:

1. **Purpose and character of the use, including whether the use is of a commercial nature or is for nonprofit educational purposes:** Courts look at how the party claiming fair use is using the copyrighted work, and are more likely to find that nonprofit educational and noncommercial uses are fair. This does not mean, however, that all nonprofit education and noncommercial uses are fair and all commercial uses are not fair; instead, courts will balance the purpose and character of the use against the other factors below. Additionally, “transformative” uses are more likely to be considered fair. Transformative uses are those that add something new, with a further purpose or different character, and do not substitute for the original use of the work.
2. **Nature of the copyrighted work:** This factor analyzes the degree to which the work that was used relates to copyright’s purpose of encouraging creative expression. Thus, using a more creative or imaginative work (such as a novel, movie, or song) is less likely to support a claim of a fair use than using a factual work (such as a technical article or news item). In addition, use of an unpublished work is less likely to be considered fair.
3. **Amount and substantiality of the portion used in relation to the copyrighted work as a whole:** Under this factor, courts look at both the quantity and quality of the copyrighted material that was used. If the use includes a large portion of the copyrighted work, fair use is less likely to be found; if the use employs only a small amount of copyrighted material, fair use is more likely. That said, some courts have found use of an entire work to be fair under certain circumstances. And in other contexts, using even a small amount of a copyrighted work was determined not to be fair because the selection was an important part—or the “heart”—of the work.
4. **Effect of the use upon the potential market for or value of the copyrighted work:** Here, courts review whether, and to what extent, the unlicensed use harms the existing or future market for the copyright owner’s original work. In assessing this factor, courts consider whether the use is hurting the current market for the original work (for example, by displacing sales of the original) and/or whether the use could cause substantial harm if it were to become widespread.

In addition to the above, other factors may also be considered by a court in weighing a fair use question, depending upon the circumstances. Courts evaluate fair use claims on a case-by-case basis, and the outcome of any given case depends on a fact-specific inquiry. This means that there is no formula to ensure that a predetermined percentage or amount of a work—or specific number of words, lines, pages, copies—may be used without permission.

<https://www.copyright.gov/fair-use/> (That page also contains a link to the Copyright Office’s Fair Use Index, which is described by the CO, in part as: The Fair Use Index tracks a variety of judicial decisions to help both lawyers and non-lawyers better understand the types of uses courts have previously determined to be fair—or not fair. The decisions span multiple federal jurisdictions, including the U.S. Supreme Court, circuit courts of appeal, and district courts. Please note that while the Index incorporates a broad selection of cases, it does not include all judicial opinions on fair use. The Copyright Office will update and expand the Index periodically.)

<sup>18</sup> Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. The full text is available at <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

<sup>19</sup> Flynn, Sean; Schirru, Luca; Palmedo, Michael; and Izquierdo, Andrés. "Research Exceptions in Comparative Copyright." (2022) PIJIP/TLS Research Paper Series no. 75. <https://digitalcommons.wcl.american.edu/research/75>

According to Jonathan Band and Jonathan Gerafi, as published on the InfoJustice website<sup>20</sup> “more than 40 countries with over one-third of the world’s population have fair use or fair dealing provisions in their copyright laws. These countries are in all regions of the world and at all levels of development.” Band and Gerafi go on to state “[t]he broad diffusion of fair use and fair dealing indicates that there is no basis for preventing the more widespread adoption of these doctrines, with the benefits their flexibility brings to authors, publishers, consumers, technology companies, libraries, museums, educational institutions, and governments. This is particularly the case considering that the copyright laws in many “civil law” countries currently allow their courts to apply a specific exception in a specific case only if the second and third steps of the Berne three-step-test are met. That is, the court may permit the use only if it determines that the use does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the rights holder. These steps are at least as abstract and difficult to apply as fair use or fair dealing.”

Band and Gerafi have published “The Fair Use/Fair Dealing Handbook”<sup>21</sup> which is an excellent resource listing numerous countries that apply Fair Dealing, and the governing statutes.

The introduction states:

Fair dealing was first developed by courts in England in the eighteenth century, and was codified in 1911. In the UK legislation, an exception to infringement was provided for fair dealing with a work for the purposes of “private study, research, criticism, review, or newspaper summary.” Fair dealing also became incorporated into copyright laws of the former British Imperial territories, now referred to as the Commonwealth countries. Over the past century, however, the fair dealing statutes have evolved in many of the Commonwealth countries. While in some countries fair dealing remains, as in the UK, restricted to the original purposes of the 1911 Act, in other countries these purposes have become a non-exclusive list of examples (*see, e.g.,* Bahamas). In still other countries, legislatures have added factors a court must consider in determining fair dealing (*see, e.g.,* Australia). Moreover, some countries have replaced the term “fair dealing” with “fair use” (*see, e.g.,* Bangladesh). Thus, the fair dealing statutes in many countries have over time increasingly resembled the fair use statute in the United States. (Additionally, judicial interpretations of fair dealing in countries such as Canada are now similar to judicial interpretations of fair use in the United States.)

---

<sup>20</sup> Band, Jonathan and Gerafi, Jonathan, “The Fair Use/Fair Dealing Handbook” at <https://infojustice.org/archives/45212> (the website hosted by the [Program on Information Justice and Intellectual Property](#) at American University Washington College of Law) borrowed pursuant to the [Creative Commons Attribution 4.0 International License](#).

<sup>21</sup> Available at <https://infojustice.org/wp-content/uploads/2023/04/Band-and-Gerafi-April-2023.pdf>. Band and Gerafi also state in the introduction: “This handbook contains the fair use and fair dealing statutes we were able to identify. [ ] Please [contact jband@policybandwidth.com](mailto:jband@policybandwidth.com) if we missed any. The handbook does not include the many implementations of the exceptions for quotations and illustration in Article 10 of the Berne Convention, which refers to “fair practice.” Fair practice under Article 10 is a distinct concept from fair use or fair dealing. The handbook also does not include the myriad specific exceptions countries have enacted in addition to fair use or fair dealing. Finally, the handbook does not contain exceptions that appear to be inspired at least in part by fair use or fair dealing, but do not employ those terms.”

Fair use in the United States is attributed to Justice Story's 1841 decision in *Folsom v. Marsh*, which was based on the English fair dealing case law. Congress codified fair use in the Copyright Act of 1976. Section 107 provides that fair use for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship and research is not an infringement of copyright. Section 107 then lists four factors that are to be included in the determination of whether the use made of a work in any particular case is a fair use. In other words, Section 107 sets forth non-exclusive purposes and non-exclusive factors for fair use. Although fair use is generally considered to be more flexible and open-ended than fair dealing, this, as discussed above, is no longer the case in many Commonwealth countries.

Countries that are not former British colonies, such as Taiwan and Korea, have also adopted fair use or fair dealing. Four former colonies, Botswana, Ghana, Lesotho, and Malawi, have replaced fair dealing with other exceptions.

The reader is also commended to the Wikipedia page on Fair Dealing.<sup>22</sup>

Machine learning developers in the US have presumed and asserted copying, ephemeral or stored, to be allowable as fair use under the Copyright Act, because the use of copyrighted work is transformative, and limited.<sup>23</sup> The situation has often been compared to Google Books's scanning of copyrighted books in *Authors Guild, Inc. v. Google, Inc.*, which was ultimately found to be fair use, because the full text of the scanned content was not made publicly available (snippets are made available), and the scanning was for a non-expressive purpose. Professor Matthew Sag in Testimony to the Senate Judiciary Committee Subcommittee on Intellectual Property on July 12<sup>th</sup>, 2023 notes that “Copyright law draws a fundamental distinction between protectable original expression, and unprotectable facts, ideas, abstractions, and functional elements. This distinction is often referred to as the idea-expression distinction or the idea-expression dichotomy. The idea-expression distinction means that copying valuable facts and ideas, or learning techniques, drawing inspiration, or emulating the general style of a copyrighted work is not infringement. In addition, the idea-expression distinction also informs the way courts apply the fair use doctrine. Reflecting the idea-expression distinction, courts have consistently held that *technical acts of copying which do not communicate an author's original expression to a new audience are fair use*. Such uses are referred to as non-expressive uses.”<sup>24</sup>

There are a number of US lawsuits disputing fair use contentions, arguing that the training of machine learning models infringed the copyright of the authors of works contained in the training data. Some of these lawsuits assert that output also infringes upon copyright.<sup>25</sup> Where

---

<sup>22</sup> [https://en.wikipedia.org/wiki/Fair\\_dealing](https://en.wikipedia.org/wiki/Fair_dealing)

<sup>23</sup> Regarding the first factor, OpenAI argues its purpose is “transformative” as opposed to “expressive” because the training process creates “a useful generative AI system.” OpenAI also contends that the third factor supports fair use because the copies are not made available to the public but are used only to train the program.

<sup>24</sup> Prof. Matthew Sag Testimony on Copyright and AI July 12, 2023 (For Distribution) (senate.gov) [https://www.judiciary.senate.gov/imo/media/doc/2023-07-12\\_pm\\_-\\_testimony\\_-\\_sag.pdf](https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_sag.pdf)

<sup>25</sup> For example, a recent lawsuit by several music publishers against AnthropicAI where it is asserted, among other claims, “Anthropic is in the business of developing, operating, selling, and licensing AI technologies. Its primary product is a series of AI models referred to as ‘Claude.’ Anthropic builds its AI models by scraping and

it is asserted that the LLM, Foundation Model, and GAI developers intentionally or negligently copied works from suspected or known infringing sources, such as in a couple of the suits by book authors, Courts might take into account whether lack of “good faith” in using the works for training weighs against the developer or user in a fair use analysis.<sup>26</sup>

---

ingesting massive amounts of text from the internet and potentially other sources, and then using that vast corpus to train its AI models and generate output based on this copied text. Included in the text that Anthropic copies to fuel its AI models are the lyrics to innumerable musical compositions for which Publishers own or control the copyrights, among countless other copyrighted works harvested from the internet. This copyrighted material is not free for the taking simply because it can be found on the internet. Anthropic has neither sought nor secured Publishers’ permission to use their valuable copyrighted works in this way.” The complaint further alleges, “As a result of Anthropic’s mass copying and ingestion of Publishers’ song lyrics, Anthropic’s AI models generate identical or nearly identical copies of those lyrics, in clear violation of Publishers’ copyrights.” *Concord Music Group, Inc. v. Anthropic PBC*, 3:23-cv-01092, (M.D. Tenn.) (<https://www.courtlistener.com/docket/67894459/1/concord-music-group-inc-v-anthropic-pbc/>)

There are also lawsuits by Getty Images against Stability AI, asserting that Stability’s use of Getty photos to train Stability constitutes copyright infringement. Or where the National Music Publisher’s Association filed suit against Twitter. Or where in June, 2023, the RIAA asked Discord to shut down a popular AI hub where, apparently, there are deepfakes of popular recording artists, and has subpoenaed Discord to provide information about the participants of that hub. Or where on June 28, 2023, a lawsuit was filed by authors against OpenAI. Sarah Silverman, often described as a comedian, is the lead plaintiff in that lawsuit, which is proposed to be certified as a class action, against OpenAI, where authors of books allege their books were used to train OpenAI’s GPT without permission based on illegal digital copies of the books from “shadow libraries” where it is alleged OpenAI is aware a great many of the books available from these “libraries” are copies of entire in-copyright books; and that GPT-4 produces detailed summaries of their books upon queries. The amount of use (substantiality) of the actual works might tend to be contrary to fair use.

(Use a search engine and search for, e.g., “Getty Images ai lawsuit.” Links: *NMPA v. Twitter* complaint filed June 14, 2023 <https://s3.documentcloud.org/documents/23848134/nmpa-v-twitter.pdf>. On the RIAA / Discord matter, see, e.g., <https://torrentfreak.com/images/RIAA-Discord-Subpoena.pdf>. On one of the lawsuits by authors against OpenAI, see, e.g., <https://www.reuters.com/legal/lawsuit-says-openai-violated-us-authors-copyrights-train-ai-chatbot-2023-06-29/>, and <https://dockets.justia.com/docket/california/candce/4:2023cv03223/414822>. *Sarah Silverman v. OpenAI, Inc.*, in the U.S. District Court for the Northern District of California, No. 3:23-cv-03223 (filed July 7, 2023) (proposed class action) ([gov.uscourts.cand.415174.1.0\\_1.pdf](http://gov.uscourts.cand.415174.1.0_1.pdf) (via [courtlistener.com](https://www.courtlistener.com/))). Some of these suits also assert claims in addition to copyright infringement claims, such as Unfair Competition, Negligence, and Unjust Enrichment. See also, Master List of lawsuits v. AI, ChatGPT, OpenAI, Microsoft, Meta, Midjourney & other AI cos., which include links to some of the complaints and court dockets <https://chatgptiseatingtheworld.com/2023/07/11/master-list-of-lawsuits-v-ai-chatgpt-openai-microsoft-meta-midjourney-other-ai-cos/>

<sup>26</sup> Legal scholar Michael W. Carroll has written in a section of a law review article that an argument against such a party is that they have no fair use rights because they are not acting in good faith. Whether good faith is, or should be, a factor in a fair use analysis is debatable. There is much more in this section of the article (III.A.2.b.) commencing at p.954 , as well as the rest of the article, of interest. Carroll, Michael W., *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful* (December 1, 2019). 53 UC Davis Law Review 893, 2019, American University, WCL Research Paper No. 2020-15, Available at SSRN: <https://ssrn.com/abstract=3531231>

Also See *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 562 (1985) (“Also relevant to the character of the use is the propriety of the defendant’s conduct. Fair use presupposes good faith and fair dealing.”) (internal citation omitted) (internal quotations omitted). 319 975 F.2d 832, 843 (Fed. Cir. 1992) (“To invoke the fair use exception, an individual must possess an authorized copy of a literary work.”). 320 108 F.3d 1119, 1122 (9th Cir. 1997) (“‘[T]he propriety of the defendant’s conduct’ is relevant to the character of the use at least to the extent that it may knowingly have exploited a purloined work for free that could have been obtained for a fee.”). 321 364 F.3d 471, 475-78, 482 (2d Cir. 2004) (reading *Harper & Row* to “direct[ ] courts to

It has been suggested that if plaintiffs succeed in some or all of the GAI copyright infringement suits, this may shift the balance of power in favor of large corporations such as Google, Microsoft and Meta who can afford to license large amounts of training data from copyright holders and leverage their own proprietary datasets of user-generated data.<sup>27</sup> Legal scholars and practitioners Mark Lemley and Bryan Casey argue that datasets are so huge that "there is no plausible option simply to license all of the (data). . . allowing (any generative training) copyright claim is tantamount to saying, not that copyright owners will get paid, but that the use won't be permitted at all."<sup>28</sup> Others disagree, with some predicting a similar

---

consider a defendant's bad faith in applying the first statutory factor" but then holding the use to be fair after such consideration). 322 510 U.S. 569, 585 n.18 (1994).

<sup>27</sup> <https://arstechnica.com/tech-policy/2023/04/stable-diffusion-copyright-lawsuits-could-be-a-legal-earthquake-for-ai/>

<sup>28</sup> "Abstract: Neural network and machine learning artificial intelligences (AIs) need comprehensive data sets to train on. Those data sets will often be composed of images, videos, audio, or text. All those things are copyrighted. Copyright law thus stands as an enormous potential obstacle to training AIs. Not only might the aggregate data sets themselves be copyrighted, but each individual image, video, and text in the data set is likely to be copyrighted too.

It's not clear that the use of these databases of copyrighted works to build self-driving cars, or to learn natural languages by analyzing the content in them, will be treated as a fair use under current law. Fair use doctrine in the last quarter century has focused on the transformation of the copyrighted work. AIs aren't transforming the databases they train on; they are using the entire database, and for a commercial purpose at that. Courts may view that as a kind of free riding they should prohibit.

In this Article, we argue that AIs should generally be able to use databases for training whether or not the contents of that database are copyrighted. There are good policy reasons to do so. And because training data sets are likely to contain millions of different works with thousands of different owners, there is no plausible option simply to license all the underlying photographs or texts for the new use. So allowing a copyright claim is tantamount to saying, not that copyright owners will get paid, but that no one will get the benefit of this new use.

There is another, deeper reason to permit such uses, one that has implications far beyond training AIs. Understanding why the use of copyrighted works by AIs should be fair actually reveals a significant issue at the heart of copyright law. Sometimes people (or machines) copy expression but they are only interested in learning the ideas conveyed by that expression. That's what is going on with training data in most cases. The AI wants photos of stop signs so it can learn to recognize stop signs, not because of whatever artistic choices you made in lighting or composing your photo. Similarly, it wants to see what you wrote to learn how words are sequenced in ordinary conversation, not because your prose is particularly expressive.

AIs are not alone in wanting just the facts. The issue arises in lots of other contexts. In *American Geophysical Union v. Texaco*, for example, the defendants were interested only in the ideas in scientific journal articles; photocopying the article was simply the most convenient way of gaining access to those ideas. Other examples include copyright disputes over software interoperability cases like *Google v. Oracle*, current disputes over copyright in state statutes and rules adopted into law, and perhaps even Bikram yoga poses and the tangled morass of cases around copyright protection for the artistic aspects of utilitarian works like clothing and bike racks. In all of these cases, copyright law is being used to target defendants who actually want something the law is not supposed to protect – the underlying ideas, facts, or functions of the work.

Copyright law should permit copying of works for non-expressive purposes. When the defendant copies a work for reasons other than to have access to the protectable expression in that work, fair use should consider under both factors one and two whether the purpose of the defendant's copying was to appropriate the plaintiff's expression or just the ideas. We don't want to allow the copyright on the creative pieces to end up controlling the unprotectable elements." Lemley, Mark A. and Casey, Bryan, *Fair Learning* (January 30, 2020). Available at SSRN: <https://ssrn.com/abstract=3528447> or <http://dx.doi.org/10.2139/ssrn.3528447>

outcome as seen in US music licensing procedures. Indeed, *we recommend statutory licensing to force technology conglomerates to pay significant sums of money to the creative industries for the use of their works while mitigating bias.*

### **Contractual Techniques for Rights-Holders Dealing with AI**

If rights-holders and website operators want to limit or prohibit scraping in the future, and cannot get scrapers to stop based on claims of copyright infringement, particularly scrapers who are or may be covered by fair use, fair dealing, or TDM exceptions, what are rights-holders to do?

In the US there were attempts by website operators to claim scraping violated the **Computer Fraud and Abuse Act (CFAA)**. The CFAA prohibits accessing a “protected computer” without authorization. We won’t get into the details of this topic, other than to say that claims of violations of the CFAA may still be viable, even if they are not being aggressively pursued, partly because of (mis)interpretations of cases such as the [hiQ Labs. V. LinkedIn](#) scraping case.<sup>29</sup> Who does get into the details is [Kieran McCarthy](#) with excellent writing on the esteemed Professor Eric Goldman’s [Technology & Marketing Law Blog](#).<sup>30</sup>

McCarthy discusses that although hiQ might have prevailed on the CFAA claims battle it lost the war. “LinkedIn’s **User Agreement unambiguously prohibits scraping and the unauthorized use of scraped data.**” LinkedIn obtained a permanent injunction and damages against hiQ Labs on that basis.

Therefore, website operators are trying to stop web scraping with **breach of contract** claims. For example, X Corp., FKA Twitter has filed multiple lawsuits against web scrapers, [including against Bright Data](#) asserting three claims: **breach of contract, tortious interference with a contract, and unjust enrichment**. McCarthy posits that plaintiffs seem confident that breach of contract claims against scrapers will prevail.

While GDMG believes that contracts might provide “solutions” to discourage scraping where website operators / publishers wish to prohibit scraping of some or all of their pages, (a) the scrapers will need to have their AI tools read the terms, etc., and abide by them, (b) rights-holders will need to consider whether copyright law (and TDM exceptions, fair dealing, and fair use) **pre-empts contract or license claims**, and (c) whether what may be called “**contractual override**” or preemption is enforceable or un-enforceable as **against public policy**. According to Jonathan Band there are 48 countries that prevent contract override, particularly where there are TDM exceptions.<sup>31</sup>

---

<sup>29</sup> At the District Court: <https://www.courtlistener.com/docket/6071320/hiq-labs-inc-v-linkedin-corporation>. At the 9<sup>th</sup> Cir. <https://www.courtlistener.com/docket/6335500/hiq-labs-inc-v-linkedin-corporation/>

<sup>30</sup> <https://blog.ericgoldman.org/archives/2023/08/web-scraping-for-me-but-not-for-thee-guest-blog-post.htm>

<sup>31</sup> “Rightsholders often distribute digital content subject to licenses that seek to override exceptions contained in national copyright laws. Recognizing that these license terms could upset their copyright law’s balance between rightsholders and users, legislators around the world have enacted clauses that invalidate license terms inconsistent with their copyright law’s exceptions. This compilation assembles the copyright override prevention clauses adopted in 48 countries over the past 30 years. It also sets forth references to contract override prevention in documents officially presented in the World Intellectual Property Organization, as well as clauses that have been proposed in various *fora* in the United States.” “For thirty years, the EU directives relating to copyright have required the nullification of license terms that override specific exceptions mandated by those

In addition, GDMG is considering whether **machine readable opt-out mechanisms** should be standardized and if utilized by rights-holders should be respected by potential scrapers, at the risk of the scrapers being subject to lawsuits, including for infringement and violations of terms and conditions. Most search engines use Robots.txt files. Profs. Lemley, et. Al, state, in part “It is worth noting that the use of a robots.txt header or other opt-out mechanism has implications for fair use also. Datasets and models like C4 (Raffel et al., 2019) and LAION-400M (Schuhmann, 2021), rely on CommonCrawl<sup>[32]</sup> data which is crawled only if users explicitly allow it through their robots.txt file. CommonCrawl is able to host a snapshot of the internet largely because of fair use arguments. As the organization’s director Gil Elbaz argues, there is a transformation into a different—not easily human-readable—format, the organization does not take a snapshot of entire webpages, and the use itself is transformative (from actively presenting content to caching content) and for the public benefit (Leetaru, 2017). In *Field v. Google, Inc.* (D. Nev. 2006), respect for the robots.txt file also was considered in the fair use assessment with the court noting that Google in good faith followed industry standards that would prevent caching (respecting disallowing crawling via a robots.txt). It is possible, then, that providing an opt-out mechanism for data creators and respecting the robots.txt opt-out mechanism will be taken into account in assessing a fair use argument, as it was in *Field v. Google, Inc.* (D. Nev. 2006).<sup>21</sup>”<sup>33</sup> See also, Google’s Developer pages for information regarding Robots.txt and not indexing.<sup>34</sup>

---

directives. The EU recognized that it would be pointless to require Member States to adopt exceptions if private parties could simply override them by contract. For example, the 2019 Directive on Copyright in the Digital Single Market renders unenforceable any contractual provision contrary to exceptions mandated under the Directive for preservation and text and data mining by cultural heritage institutions.” Band, Jonathan. "Protecting User Rights Against Contract Override," PIJIP/TLS Research Paper Series no. 97. <https://digitalcommons.wcl.american.edu/research/97>

The reader is commended to the above paper, listing the countries and states and statutes. The reader is also commended to International Federation of Library Associations and Institutions, Protecting Exceptions Against Contract Override, [https://www.ifla.org/wpcontent/uploads/2019/05/assets/hq/topics/exceptions-limitations/documents/contract\\_override\\_article.pdf](https://www.ifla.org/wpcontent/uploads/2019/05/assets/hq/topics/exceptions-limitations/documents/contract_override_article.pdf) (2019).

<sup>32</sup> This footnote and its text are by GDMG, not in the quoted original: Common Crawl is a California non-profit established by founder and Chairman Gil Elbaz. According to his website, Elbaz sold his startup Applied Semantics to Google and it became AdSense, a primary money maker for Google, and he became, and for a time was, Google’s Director of Engineering. <https://commoncrawl.org/team/gil-elbaz-chairman>

<sup>33</sup> Henderson, Peter and Li, Xuechen and Jurafsky, Dan and Hashimoto, Tatsunori and Lemley, Mark A. and Liang, Percy, Foundation Models and Fair Use (March 27, 2023). Stanford Law and Economics Olin Working Paper No. 584, Available at SSRN: <https://ssrn.com/abstract=4404340> or <http://dx.doi.org/10.2139/ssrn.4404340> ©2023 Henderson, Li, Jurafsky, Hashimoto, Lemley, & Liang. License: CC-BY 4.0, see <https://creativecommons.org/licenses/by/4.0/>.

<sup>34</sup> <https://developers.google.com/search/docs/crawling-indexing/robots/intro>. Google has posted this on that page: “**Understand the limitations of a robots.txt file.**

Before you create or edit a robots.txt file, you should know the limits of this URL blocking method. Depending on your goals and situation, you might want to consider other mechanisms to ensure your URLs are not findable on the web.

- **robots.txt rules may not be supported by all search engines.**  
The instructions in robots.txt files cannot enforce crawler behavior to your site; it’s up to the crawler to obey them. While Googlebot and other respectable web crawlers obey the instructions in a robots.txt file, other crawlers might not. Therefore, if you want to keep information secure from web crawlers, it’s better to use other blocking methods, such as [password-protecting private files on your server](#).
- **Different crawlers interpret syntax differently.**  
Although respectable web crawlers follow the rules in a robots.txt file, each crawler might interpret the rules differently. You should know the [proper syntax](#) for addressing different web crawlers as some might not understand certain instructions.

Below is an excerpt of how OpenAI is or was addressing the indexing challenge:

### **OpenAI Introduces GPTBot: A Web Crawler Designed to Scrape Data from the Entire Internet Automatically**

“OpenAI has responded to privacy and intellectual property concerns arising from data collection on public websites by introducing a new web crawler tool called GPTBot. This technology aims to gather public web data transparently and utilize it for training their AI models, all under the umbrella of OpenAI's banner. GPTBot's user agent aims to amass data that will contribute to refining future AI models. During this process, GPTBot will omit sources that necessitate payment. However, it's important to note that some collected data may inadvertently contain identifiable information or text, violating OpenAI's policies. OpenAI recognizes the need to provide website administrators with options concerning GPTBot's platform access. Granting access is perceived as a collaboration in improving the precision of AI models, ultimately enhancing their capabilities and reinforcing security measures. Conversely, OpenAI has outlined a procedure for those who prefer not to include their websites in GPTBot's data collection efforts. This guidance includes incorporating GPTBot directives into the website's robots.txt file and configuring its access to specific content segments.”<sup>35</sup>

According to Axios as reported on August 31, 2023, Nearly 20% of the top 1000 websites in the world are blocking certain crawler bots for certain data curators that gather web data for AI services, according to data from [Originality.AI](https://www.originality.ai/), an AI content detector.<sup>36</sup>

The following post describes some of the problems with opt-out schemes, though. “It's clear that if generative AI companies want to do right by artists, their current business model is

- **A page that's disallowed in robots.txt can still be indexed if linked from other sites.** While Google won't crawl or index the content blocked by a robots.txt file, we might still find and index a disallowed URL if it is linked from other places on the web. As a result, the URL address and, potentially, other publicly available information such as anchor text in links to the page can still appear in Google search results. To properly prevent your URL from appearing in Google search results, [password-protect the files on your server](#), [use the noindex meta tag or response header](#), or remove the page entirely.”

<sup>35</sup> <https://www.marktechpost.com/2023/08/10/openai-introduces-gptbot-a-web-crawler-designed-to-scrape-data-from-the-entire-internet-automatically/>

And see <https://platform.openai.com/docs/gptbot>, which contains this, for example:

#### Disallowing GPTBot

To disallow GPTBot to access your site you can add the GPTBot to your site's robots.txt:

```
User-agent: GPTBot
Disallow: /
```

#### Customize GPTBot access

To allow GPTBot to access only parts of your site you can add the GPTBot token to your site's robots.txt like this:

```
1 User-agent: GPTBot
2 Allow: /directory-1/
3 Disallow: /directory-2/
```

<sup>36</sup> <https://www.axios.com/2023/08/31/major-websites-are-blocking-ai-crawlers-from-accessing-their-content> and Originality.AI, which updated its data and reports that as of September 22, 2023, approximately 25% of those websites are blocking certain data curators from scraping their sites. Originality.AI has some interesting information.

impractical. Creating opt-outs and other stop-gap interventions will only go so far, since they do nothing to change companies' business models or challenge prevailing labor conditions.”<sup>37</sup>

Further, Henderson, et al. point out in “Foundation Models and Fair Use” that “It is possible that some strategies could be pursued that would compensate data creators even when model training meets existing fair use standards, but these should be handled with care to avoid an alternative outcome that aggregates power in other undesirable ways. For example, forcing licensing mechanisms or opt-in approaches for all data could consolidate power in those companies that already have licenses to enormous amounts of data, like YouTube or Facebook. Or they could create powerful intermediaries that aggregate data licenses without actually sufficiently compensating data creators.”<sup>38</sup>

GDMG is considering whether machine readable mechanisms that set forth licensing terms or conveys a message that the rights-holder is amenable to negotiating licensing terms, should be standardized. (See, e.g., <https://site.spawning.ai/spawning-ai-txt>.)

Risks of requiring opt-out, opt-in, and licensing, and respecting “no scraping” tech includes that GAI tools might be substandard, **biased**, or discriminatory, if the universe of materials available to train is constrained, or materials and data are not high quality. In music, certain genres and languages may be under-represented in the training set if these mechanisms are standardized and it is required that they be respected. (See, the section titled **Mitigate against Bias**, and accompanying footnotes.) In addition, as GAI tools such as ChatGPT evolve and integrate with the internet and tools and apps, such as Word, and Gmail, more intimately, these tools could serve a role similar to that of a search engine. By providing users with direct links or references from web sources, ChatGPT can direct what would be presumed to be significant traffic to those sites. When we have used ChatGPT and Bard we have clicked through to some of the websites referenced in the output. So, if a website blocks a data curator from scraping that site, that site's content may not be among the recommended or referenced sources. Just as blocking Google would prevent a website from appearing in one of the world's most popular search engines, blocking scraping might mean missing out on visitors to web sites. (Also, see, e.g., the stance of Japan that copyright does not apply in using works to create databases to train GAI tools, so that it and companies in Japan may have advanced GAS tools <https://technomancers.ai/japan-goes-all-in-copyright-doesnt-apply-to-ai-training/>) and the section titled **Harmonization**.)

There is also risk associated with GAI tools training on data and content created in whole or in part by GAI tools. Recent research papers have discussed potential GAI “disorders” that are emerging as the technology is more widely deployed and used. "**Model collapse**," "**Model Autophagy Disorder**," and "**Habsburg AI**" are what some researcher call what happens to GAI models trained and retrained using data produced by other GAIs rather than

---

<sup>37</sup> Sayash Kapoor and Arvind Narayanan. Artists can now opt out of generative ai. it's not enough. <https://www.aisnakeoil.com/p/artists-can-now-opt-out-of-generative>

<sup>38</sup> See FN<sup>33</sup>.

humans.<sup>39</sup> We stand for the proposition that technologies must be created and implemented that prevent the scraping of GAI generated content and the use of GAI generated content to train GAI tools.

Furthermore, to the extent entities failed to attach or include meta-data associated with each Work, such as source and **Copyright Management Information** under the Digital Millennium Copyright Act of 1998 (“DMCA”), such failures may make it exceedingly difficult for such entities to reverse engineer or determine who the rights-holders are in order to either seek permission or pay negotiated or compulsory<sup>40</sup> licensing fees to the extent such information is missing or incomplete. *Now is a good time to consider an open source blockchain based solution to provide copyright management information for GAI companies to ingest and from which to provide provenance information to enable business models and compliance with existing and future laws and regulations, including Federal Trade Commission truth in advertising laws.*

### **Conclusion on Ingestion**

Based on the above considerations of use of copyright, fair use, TDM exceptions, Computer Fraud and Abuse, User Agreements, DMCA and Copyright Management, **GDMG believes that statutory licensing from the rightsholder is or should be required for ingestion** or mathematical representation of intellectual property works in all GAI tools where commercial use is probable beyond the limited non-commercial uses provided for in Fair use and Text and Data Mining exceptions. In the alternative, where there is a direct licensing instead of statutory licensing paradigm, we recommend there should be a corresponding exemption from direct licensing requirements for ingestion if **standard technical measures** have been taken to respect the opt-in or opt-out choices of webpages and data-set holders as well as measures to prevent certain ingestions and to block the creation of certain outputs. This is similar to the safe harbor and “standard technical measures” concepts in the DMCA in the United States, provided that at present we highly recommend that legislators delegate authority to an agency to continually define what standard technical measures and what other requirements of safe harbor are as this terminology under the DMCA was never well defined in law or courts and has led to hundreds of billions of dollars in unintended value transfer from copyright owners to technology conglomerates.

## **2. INTELLECTUAL PROPERTY ISSUES WITH REGARD TO THE PROMPTING AND OUTPUT OF FOUNDATION MODELS & GAI TOOLS:**

We’ve spent some time on issues regarding the ingestion of content. We will now focus on the use and output side of GAI tools.

### **When is there Copyright Infringement by GAI Created Outputs?**

It is a complex issue regarding whether and when a GAI generated output in fact contains elements of another rights-holder’s in-copyright work or is substantially similar. If either of these conditions are true, such output would likely infringe upon the rights-holder’s copyright, but for fair use, fair dealing, or a permitted use.

<sup>39</sup> See, e.g., <https://www.axios.com/2023/08/28/ai-content-flood-model-collapse>

<sup>40</sup> See, e.g., [https://en.wikipedia.org/wiki/Mechanical\\_license](https://en.wikipedia.org/wiki/Mechanical_license) regarding compulsory mechanical licenses in connection with new sound recordings or “covers” of musical compositions.

### Copyrightability of GAI Works

Authorship requires creative human contribution. Works consisting of both human-authored and GAI generated material may be copyrightable. We suggest that the more creativity and details utilized in prompts, mechanisms of control, and illustrative inputs, and selection of the output, the more a work will qualify for copyright registration. We recommend that the copyright office of the United States develop guidelines for how human creativity and intervention is measured when determining if a work is human or machine authored. We believe these standards should go beyond the type of ‘self-certification’ that exists in Lanham Act and Federal Trademark applications for marks related to performers. The solution probably involves a programmatic connection using **industry-government defined communication standards** between GAI companies and the copyright office. This application programming interface of sorts would enable authentication of the prompt engineering involved in the output. This is another application for an industry supported open-source blockchain based solution for provenance information in content.

Furthermore, in instances where a GAI tool “creates” a work 100% based on the work of an author or rights holder, the output should be entitled to copyright protection as a work derivative of the original. For example, if a GAI tool has ingested all of the musical compositions of Artist A, and all of the sound recordings of Artist A, and the rights holder(s) of those compositions and sound recordings instruct a GAI tool to create a musical composition and sound recording solely in the styles of Artist A, the output should be entitled to copyright protection inuring to the benefit of the Artist whose NILVS is reproduced, as well as the rights-holders of the musical compositions and sound recordings used to create the output(s). Perhaps there should be some co-authorship rights to the producers (prompt engineers), mixers, etc.

If a GAI tool’s output resembles (is substantially similar to) copyrighted works and affects the market for the copyrighted work, it is our opinion that fair use may not and should not apply. **Standard Technical Measures** should co-evolve, potentially introducing **Safe Harbors** when strong technical tools are employed to allow copyright holders to manage their rights.

### Right of Publicity and Privacy, Impersonations, and Deep Fakes

There are a few ways in which people’s name, image, likeness, voice, style, persona, or performance (what we are calling NILVS) can be used in GAI: (1) NILVS, including videos, and sound recordings, of humans can be used as training data; (2) people’s names and other identifying data, such as videos, images, and sound recordings can be used as prompts; (3) outputs of GAI can include, replicate, or evoke the NILVS of identifiable humans. Some of the issues of concern regarding these include threats to: the right of humans to control how our identities are used; for some, their livelihood; democracy, because GAI might produce deepfakes of politicians giving speeches or doing things they did not say or do, or of influencers purportedly endorsing a candidate or policy position.

Stated generally, a right of publicity is currently the right of humans to control the **commercial**<sup>41</sup> use of their identity, for example, when an individual’s identity is used in an

---

<sup>41</sup> Commercial speech is “speech which does no more than propose a commercial transaction.” *Bolger v. Youngs Drug Prods. Corp.*, 463 U.S. 60, 66 (1983) (internal quotation marks omitted). “If speech is not ‘purely commercial’—that is, if it does more than propose a commercial transaction—then it is entitled to full First Amendment protection.” *Mattel, Inc. v. MCA Records, Inc.*, 296 F. 3d 894, 906 (9th Cir. 2002).

advertisement or on products, goods, or services. Identity has traditionally comprised an individual's name, likeness, voice, and signature. It is important to note that the First Amendment prohibits restraining the use of an individual's identity in "non-commercial" expressive speech, such as in news reporting, commentary, entertainment, works of fiction or nonfiction, or in advertising that is incidental to such uses, even if such uses are for profit.<sup>42</sup> We will look at how First Amendment issues are addressed in the proposed Federal NO FAKES Act discussed below.

Although there is currently no federal right of publicity, it is reported that a majority of states currently recognize a statutory or common law right of publicity, or both, through a patchwork or hodgepodge of laws.<sup>43</sup> As under the Lanham Act, discussed below, the laws of almost all states only protect those who are famous or publicly recognizable or whose identities have commercial value.<sup>44</sup> Right of publicity claims do not require that there is a likelihood of confusion, so a defendant who violates the right could not avoid liability by adding a disclaimer to GAI output. Some states recognize a right through the tort of invasion of privacy by appropriation under which a plaintiff can recover for a defendant's unpermitted use of plaintiff's identity with damage to plaintiff's dignitary interests and peace of mind. Other laws could potentially be invoked in connection with GAI generated outputs that include the name or likeness, including vocal likeness, of a particular human, including defamation or false light (e.g., if a digital replica falsely depicts a human doing something they did not actually do or say, and where the use injures the individual's reputation or is highly offensive to them),<sup>45</sup> fraud, or trademark infringement. Imitating an individual's style usually does not receive protection, including under copyright law.

---

<sup>42</sup> See Restatement (Third) of Unfair Competition, § 47. See, also, e.g., (See Robert C. Post and Jennifer E. Rothman, [The First Amendment and the Right\(s\) of Publicity](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3553946), 130 Yale L.J. 86 (2020) [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3553946](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3553946))

<sup>43</sup> For a survey of the US states and foreign jurisdictions that protect the right of see International Trademark Association, Right of Publicity Committee, Right of Publicity State of the Law Survey (2019), at: [https://www.inta.org/wp-content/uploads/public-files/advocacy/committeereports/INTA\\_2019\\_rop\\_survey.pdf](https://www.inta.org/wp-content/uploads/public-files/advocacy/committeereports/INTA_2019_rop_survey.pdf) Prof. Rothman has surveyed the laws and case law of various states and federal courts. See, e.g., this page with regard to California, the home of GDMG: [California – Roadmap to the Right of Publicity \(rightofpublicityroadmap.com\)](https://rightofpublicityroadmap.com/state_page/california/) ([https://rightofpublicityroadmap.com/state\\_page/california/](https://rightofpublicityroadmap.com/state_page/california/))

<sup>44</sup> An exception to this is New York's right of publicity statute, which provides a cause of action against anyone who "discloses, disseminates or publishes sexually explicit material[s]" that includes "computer-generated nude body parts as the nude body parts of the depicted individual or the depicted individual engaging in sexual conduct ... in which the depicted individual did not engage." NY S.B. 5959, available at <https://legislation.nysenate.gov/pdf/bills/2019/S5959D>.

<sup>45</sup> Indeed, as of this writing there is a defamation lawsuit pending against OpenAI, the first to be reported. Talk-radio host [Mark Walters](https://www.courtlistener.com/docket/67617826/walters-v-openai-llc/) filed a lawsuit against OpenAI in Georgia's state court, which has been removed to federal court. Walters contends in the amended complaint ([Amended Complaint – #30 in Walters v. OpenAI, L.L.C. \(N.D. Ga., 1:23-cv-03122\) – CourtListener.com](https://www.courtlistener.com/docket/67617826/walters-v-openai-llc/)) that ChatGPT, responding to a journalist's prompt to summarize the allegations in a different [lawsuit's complaint](https://www.courtlistener.com/docket/67617826/walters-v-openai-llc/), hallucinated and falsely wrote that Walters was a defendant "accused of defrauding and embezzling funds from" the lead plaintiff, a non-profit organization. The journalist asked ChatGPT to provide a copy of that complaint, and the output provided is alleged to be "a complete fabrication and bears no resemblance to the actual complaint, including an erroneous case number." Walters was not even a party in that lawsuit. Walters's [complaint](https://www.courtlistener.com/docket/67617826/walters-v-openai-llc/) states "[E]very statement of fact in the summary pertaining to Walters is false." The journalist never published the false information about Walters, which brings into question what Walters's damages might be. *Walters v. OpenAI, L.L.C.*, 1:23-cv-03122, (N.D. Ga.) (<https://www.courtlistener.com/docket/67617826/walters-v-openai-llc/>).

The Lanham Act<sup>46</sup> is a possible applicable US federal law in this arena that applies trademark protection from false designation of origin. For performers and others, the typical Lanham Act cause of action is false endorsement. However, currently, it only applies to individual's deemed "famous" or "publicly recognizable" or who have commercial value, under the Second Circuit's Carmen Elektra decision.<sup>47</sup> Trademark and unfair competition laws, both under the federal Lanham Act and state laws, protect against unauthorized uses of a person's identity if the use suggests confusion as to the person's endorsement or participation in the newly generated work. Although Federal Trademark registration is not required for a Lanham Act false endorsement claim, registration may be limited to a particular aspect of a celebrity's persona, not a general reference to their identity, and style and voice are likely not protected.

As implicated in the discussion at footnote <sup>45</sup> about the Walters defamation lawsuit, who does an aggrieved rights-holder sue? The GAI model deployer, such as OpenAI in connection with ChatGPT? The GAI model user? The GAI output user who distributes, displays, exhibits, publishers, performs, etc., the allegedly libelous GAI output?

Although there are various state laws regarding certain identity rights, but principally only in connection with commercial uses, the First Amendment prohibits restraining the use of an individual's identity in "non-commercial" expressive speech, and Lanham Act and other claims or causes of action appear to only protect those who are famous, publicly recognizable, or have commercial value (at least in the Second Circuit), and there are no US federal laws, or laws of other countries protecting against deepfakes, impersonations, or unauthorized appropriation of NILVS. There are calls for such laws. There are legitimate points by both opponents and proponents for the need of federal law to protect against the unauthorized uses of an individual's identity, and what the details should and should not be. We believe that it is imperative that well considered laws be passed in the United States on the federal level, and harmonized way around the world. A fair attempt at that has been started.

At a hearing before the U.S. Senate Judiciary Committee's Subcommittee on Intellectual Property on the July 12, 2023, which opened with Senator Coons playing a recording of a GAI-generated song about GAI set to the tune of *New York, New York* in the vocal style of Frank Sinatra. After this entertaining opening there were questions and statements by Senators, discussion, and testimony about fakes and rights of publicity. There were follow-up

---

See also this blog post by Professor Eugene Volokh regarding the case: [First \(?\) Libel-by-AI \(ChatGPT\) Lawsuit Filed \(reason.com\)](https://reason.com/volokh/2023/06/06/first-ai-libel-lawsuit-filed/printer/) \* <https://reason.com/volokh/2023/06/06/first-ai-libel-lawsuit-filed/printer/>, and Prof. Volokh's paper [Large Libel Models? Liability for AI Output \(ucla.edu\)](https://www2.law.ucla.edu/volokh/ailibel.pdf) (<https://www2.law.ucla.edu/volokh/ailibel.pdf>), wherein, among other things, an argument is made that ChatpGPT is likely not shielded from liability by Section 230 of the Communications Decency Act. 47 U.S.C. Sec. 230 states that, "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." Others argue Sec. 230 does shield ChatGPT from liability. See, e.g., Jess Miers, Yes, Section 230 Should Protect ChatGPT and Other Generative AI Tools, TECHDIRT, Mar. 17, 2023, 11:59 am, <https://perma.cc/ZH73-N3XA>.

<sup>46</sup> <https://www.law.cornell.edu/uscode/text/15/1125>

<sup>47</sup> <https://law.justia.com/cases/federal/appellate-courts/ca2/19-235/19-235-2021-02-09.html>

written questions and testimony.<sup>48</sup> On October 12, 2023, several U.S. Senators released a bi-partisan discussion draft of what they are calling the Nurture Originals, Foster Art, and Keep Entertainment Safe (NO FAKES) Act, intended to protect the voice and visual likenesses of humans from unauthorized uses through GAI.<sup>49</sup> Furthermore, the Copyright Office published a Notice of Inquiry to collect information and views relevant to the copyright law and policy issues raised by recent advances in generative AI.<sup>50</sup> So far there are more than ten thousand submissions, and the deadline for reply comments has not yet passed. Questions 30 to 33 ask about publicity rights.<sup>51</sup> And on September 29, 2023, the Congressional Research Service published a bi-partisan report on the issue.<sup>52</sup> Clearly, there is great concern about the misuse of GAI to create unauthorized uses and fakes that include elements of an individual’s identity.

The NO FAKES act would create a digital replication property right in living and dead humans to “authorize the use of the image, voice or visual likeness of the individual in a digital replica.” In the case of the dead, this property right would be held by their executor, heir, assign, or devisee. Humans may license their rights to a third party, such as a record label, movie studio, etc. We are anticipating proposed federal legislation may be fast tracked.

Like many others, we believe many of the proposals in the NO FAKES act draft are good, some are confusing or ambiguous, and there’s room for improvement. We do not provide fully detailed analysis in this paper. Congressional oral and written testimony is made public, as are comments to Notices of Inquiry, but letters and submissions to Senators and Representatives are not typically made public except by the author or submitter of same. We don’t know what feedback the Senators proposing the draft of the NO FAKES act have received from experts on various aspects of the issues raised by the NO FAKES draft.

---

<sup>48</sup> Video and written submissions may be accessed here: <https://www.judiciary.senate.gov/artificial-intelligence-and-intellectual-property-part-ii-copyright>. The video starts at about the 19:00 minute mark, with the GAI-generated Frank Sinatra-like song starting shortly after that.

<sup>49</sup> The one pager is here: [NO FAKES Act one pager \(senate.gov\)](https://www.blackburn.senate.gov/services/files/D33836B6-7A4B-4F09-B542-E3FC255D411F) (<https://www.blackburn.senate.gov/services/files/D33836B6-7A4B-4F09-B542-E3FC255D411F>)  
The discussion draft is here: [EHF23968 \(senate.gov\)](https://www.coons.senate.gov/imo/media/doc/no_fakes_act_draft_text.pdf) ([https://www.coons.senate.gov/imo/media/doc/no\\_fakes\\_act\\_draft\\_text.pdf](https://www.coons.senate.gov/imo/media/doc/no_fakes_act_draft_text.pdf))  
Separate legislation to Protect [Federal] Elections from Deceptive AI has been drafted (See, Text of S. 2770: Protect Elections from Deceptive AI Act (Introduced version) - GovTrack.us (<https://www.govtrack.us/congress/bills/118/s2770/text>) and H.R. 3044 - REAL Political Advertisements Act (<https://www.congress.gov/bill/118th-congress/house-bill/3044?q=%7B%22search%22%3A%22hr3044%22%7D&s=1&r=1>))

<sup>50</sup> See, <https://www.regulations.gov/docket/COLC-2023-0006>.

<sup>51</sup> Additional Questions About Issues Related to Copyright:

- What legal rights, if any, currently apply to AI-generated material that features the name or likeness, including vocal likeness, of a particular person?
- Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material? If so, should it preempt state laws or set a ceiling or floor for state law protections? What should be the contours of such a right?
- Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works “in the style of” a specific artist)? Who should be eligible for such protection? What form should it take?

<sup>52</sup> Artificial Intelligence Prompts Renewed Consideration of a Federal Right of Publicity. [LSB11052 \(congress.gov\)](https://www.congress.gov) (<https://crsreports.congress.gov/product/pdf/LSB/LSB11052>)

However, we will break the proposal down a bit by looking at parts of the draft, touch upon some of the thoughts of others that we are aware of,<sup>53</sup> and provide some of our current thinking and advocacy.

As stated above, the created right is designated as a property right.

The right would, essentially, prohibit anyone who is not the rights holder from: (A) producing a digital replica without the consent of the applicable human or rights holder. (B) from publishing, distributing or transmitting, or otherwise making available to the public, an unauthorized digital replica, if the person (or entity) engaging in any of the foregoing has knowledge that the digital replica was not authorized by the applicable human or rights holder. So, the production portion of subsection (A) seems to be a strict liability provision, and the other activities (subsection (B)) would require knowledge that the rights holder did not authorize the digital replica.

Who is the producer? The GAI model deployer, such as OpenAI in connection with ChatGPT? The GAI model user? Both? Who is publishing, distributing or transmitting, or otherwise making available to the public, an unauthorized digital replica? The GAI model deployer? The GAI model user? The GAI output user who distributes, displays, exhibits, publishes, performs, etc., the GAI output? All three?

We think it is appropriate that the rights cover humans regardless of fame, whether they were publicly recognizable, or had / have commercial value. Any federal legislation in this area should make sure that everyone benefits from its protections, not solely famous and commercially successful people. Many ordinary people have their names, likenesses, and voices used without their permission in ways that cause significant harm, including reputational and commercial injuries. There should be no requirement to have a commercially valuable identity to bring a claim. We note, however, that regardless of whether an individual is famous or not, any right should probably address “doppelganger” problems, that is,

---

<sup>53</sup> For example, we are aware of some of the comments to the CO NOI, as well as comments that have been made public by advocacy groups and scholars, some of which predate the release of the draft of the NO FAKES Act. So far many, if not most, of the advocacy groups, including some we are a member of, have said with regard to the draft of the NO FAKES Act, something to the effect they are looking forward to working with Congress. Professor Jennifer E. Rothman and the Electronic Frontier Foundation have been critical of the need for and dangers of a federal right of publicity law. We suggest a review of the following, cherry picked, and in no particular order, with the knowledge that there are many more comments addressing the issue of a federal right of publicity law:

1. Comment from American Association of Independent Music (A2IM) and the Recording Industry Association of America, Inc. (RIAA) (<https://www.regulations.gov/comment/COLC-2023-0006-8833>)
2. Comment from Rothman, Professor Jennifer (<https://www.regulations.gov/comment/COLC-2023-0006-8229>)
3. A Broad Federal Publicity Right Is a Risky Answer to Generative AI Problems, EFF, McSherry, Corryne (<https://www.eff.org/deeplinks/2023/07/broad-federal-publicity-right-risky-answer-generative-ai-problems>)
4. Comment from Electronic Frontier Foundation (<https://www.regulations.gov/comment/COLC-2023-0006-8949>)
5. Comment from Donaldson Callif Perez LLP (<https://www.regulations.gov/comment/COLC-2023-0006-9005>)
6. Comment from Motion Picture Association, Inc. (<https://www.regulations.gov/comment/COLC-2023-0006-8970>)

consideration should be given to dealing with situations in which a GAI output by chance looks or sounds like another individual. Such instances should not give rise to strict liability, especially if the output was not produced or used for commercial purposes.<sup>54</sup>

### **Licensing And Descendibility / Assignment & Transferability**

The property right would be “descendible and licensable in whole or in part, by the individual to whom the right applies.” While we agree that such rights should be licensable, and the rights, including licensing rights, should descend, with concerns and thoughts including length of time, discussed further below, we differentiate licensing from assignment and transferability. We believe that any right created by federal legislation should not be transferable from humans. Creating a transferable right would strip individuals of the protections that the proposed legislation is supposed to provide. Allowing another person or entity to own a living human being’s name, likeness, voice, or other indicia of a person’s identity in perpetuity, or even for 70 years after that individual’s death, strikes us as impinging upon an individual’s constitutional rights and liberty, and should not be allowed.

Further, any license to use another person’s identity should only authorize specific uses, types of performances, or set of uses and performances, over which the individual rights-holder (or descende) has some ongoing reasonable control or right of refusal over the words and actions of identifiable aspects of their digital replicas. Perhaps something akin to the moral rights in copyright provided for by certain countries, but more protective.<sup>55</sup> For example, would Nobel Laureate and Holocaust camp survivor Elie Weisel<sup>56</sup> want anything identifiable about him associated with anything produced and distributed by a white-supremacist organization where he is made to appear to be a member of that organization? We think not. Furthermore, we do not want licensees to generate digital replication performances that appear to be by an individual forever and in unknown contexts as if they are saying or singing words they never said nor agreed. To allow this might also undermining trust in authentic communications because people will not know what to trust, as if that isn’t already problem enough.

According to Professor Jennifer Rothman at the University of Pennsylvania School of Law the descendibility provision tracks language in state publicity laws that do not require any commercial exploitation during a person’s lifetime. Prof. Rothman points out the draft seems to create a postmortem right in every human that has ever lived regardless of whether they were performers, or had or have commercial value, and grants rights in every dead person dating back 70 years from enactment. The provision is not limited to those who died in the United States so seems to apply to those who died anywhere in the world.<sup>57</sup>

---

<sup>54</sup> Some celebrities look like other celebrities. See, e.g., ["Lookalikes" A list of actors and celebrities that look alike - IMDb](https://www.imdb.com/list/ls033198687/) (https://www.imdb.com/list/ls033198687/). There are also businesses that act as agencies to provide celebrity look-alikes for all sorts of purposes.

<sup>55</sup> [Moral rights - Wikipedia](https://en.wikipedia.org/wiki/Moral_rights#cite_note-:0-9) (https://en.wikipedia.org/wiki/Moral\_rights#cite\_note-:0-9)

<sup>56</sup> See, e.g., [Elie Wiesel - Wikipedia](https://en.wikipedia.org/wiki/Elie_Wiesel) (https://en.wikipedia.org/wiki/Elie\_Wiesel)

<sup>57</sup> See, e.g., The comments of Prof. Rothman to the CO NOI: Jennifer E. Rothman, Summary and Analysis of Proposed NO FAKES ACT of 2023, Discussion Draft Dated October 11, 2023, accessible at [https://rightofpublicityroadmap.com/wp-content/uploads/2023/10/Prof-Rothman-Comments-to-Copyright-Office-on-Right-of-Publicity-and-AI\\_October-2023.pdf](https://rightofpublicityroadmap.com/wp-content/uploads/2023/10/Prof-Rothman-Comments-to-Copyright-Office-on-Right-of-Publicity-and-AI_October-2023.pdf) and [Regulations.gov](https://www.regulations.gov/comment/COLC-2023-0006-8229) (https://www.regulations.gov/comment/COLC-2023-0006-8229)

In terms of licensing by the applicable individual during their lifetime, or their executors, heirs, assigns, or devisees of the applicable individual for a period of 70 years after the death of the individual, we have a few thoughts. First, it should be clarified, the rights descend to not only the first set of descendees, but however many there may properly be for whatever the agreed upon period of time. What if an individual bequeaths their rights to someone who then lives only a short period of time after the applicable individual has died? What happens to the rights of publicity? Do they disappear and no further heirs enjoy the right? Does the licensee lose the right for the remainder of the term? We won't even get into whether post-mortem rights create perverse incentives for descendees to commercialize dead celebrities even if they and their loved ones do not want to, or believe that the dead celebrity would not want them to, particularly in certain contexts, because of how such a right of publicity would be treated under current estate tax laws, but Prof. Rothman does mention that in her comments to the Copyright Office NOI and cites to her webpage post about the matter involving the Michael Jackson estate.<sup>58</sup>

Although the discussion draft provides that a license is “valid only if (i) the applicable individual was represented by counsel in the transaction and the assignment agreement was in writing; or (ii) the licensing of the right covered by the assignment is governed by a collective bargaining agreement” we think mere representation by counsel is inadequate to protect the rights holder (as distinguished by the licensee) against unequal bargaining power. In law in action an attorney can only explain that the proposed license term is very long (if it is), and the rights holder might be licensing away their rights for relatively little present value compared to what might be significantly greater value in the future, and, therefore, the license period should be shorter or more limited, or there should be a right of renegotiation, or there should be renewal option periods and terms, and so forth. But a starving artist faced with very little negotiating leverage might not take the attorney's advice. Making publicity rights or other identity-based rights licensable or transferable might present risk to student and newly professional athletes, nascent and aspiring actors, recordings artists, influencers, etc., who might be pressured to grant such rights in perpetuity, or whose parents or guardians may do so when children are minors. Licensing for long periods, and transferability also threatens individual who may unwittingly sign over those rights as part of online terms-of-service that they click approval of without even reading. Perhaps the requirement of attorney review might mitigate against this, but there's no such assurance.

There have been suggestions that the term of a license should not exceed seven years, which is an outer limit in personal services contracts. We look at this suggestion weighed against a termination right similar to the termination right under the Copyright Act, weighed against what is in the current draft. A termination right under the Copyright Act permits authors or their heirs, under certain circumstances, to terminate the exclusive or nonexclusive grant of a transfer or license of an author's copyright in a work or of any right under a copyright. These provisions are intended to protect authors and their heirs against early agreements that don't recognize the current or future economic value of their works by giving them an opportunity to share in the later economic success of their works by allowing authors or their heirs, during particular periods of time long after the original grant (typically about 35 years), to regain the previously granted copyright or copyright rights. Note that grants made via a will or involving

---

<sup>58</sup> Jennifer E. Rothman, *Mixed Victory for Jackson Estate in Tax Court*, at [https://rightofpublicityroadmap.com/news\\_commentary/mixed-victory-jackson-estate-tax-court/](https://rightofpublicityroadmap.com/news_commentary/mixed-victory-jackson-estate-tax-court/) (discussing taxation of state postmortem publicity rights) (May 18, 2021)

a work made for hire may not be terminated under these provisions.<sup>59</sup> Of course there were vociferous arguments around the creation of termination rights when such rights were being considered, some of which are not that much different than applicable to the draft of the NO FAKES act.

A creator / performer / author can typically create additional copyrightable works in their lifetime, and unless the author extends or enters into another agreement pertaining to their new or as yet to be created works, works created by the author after a 7 year personal contract period would belong to the author rather than being treated as a work made for hire or assigned to the assignee under a personal services contract, similar to contracts many recording companies have with authors / performers. An individual typically only has one set of identifying features and can't create and economically capitalize on more identifying features. And so, an individual with little bargaining leverage who signs away some or all of their digital replication rights in many or all fields for many years just doesn't seem right.

Further, any licenses involving children under a certain age, perhaps 13, should expire when they turn 18, and regardless, any proposed license with a minor should first be reviewed by a court given the potential for parents, guardians, foster parents, etc., to sometimes exploit the income potential of a child and not act in the best interests of the minor, and to blind side them.<sup>60</sup> Additionally, income earned from such licensing should be held in a trust for them, until they reach the age of majority, similar to what occurs for child actors in California under what is known as Coogan's Law.<sup>61</sup>

On the other side of the coin a licensee might take present and future economic risk, and incur lost opportunity costs, to license and exploit digital replication publicity rights. These might include paying the licensor for the rights, and thereafter incurring cost for the production of content containing identifying features of the individual, and marketing, distributing, and otherwise exploiting such content, and protecting against that content and the identifying features of the individual from being infringed upon.

There's much more complexity to termination rights under copyright than the preceding might lead one to infer, but isn't there always some devil in the details? Nonetheless, there would likely be complexity in details regarding termination rights for licenses of rights of publicity, but that doesn't mean there should not be termination rights.

### **Damages and Liability**

Any legislation should include statutory damages to protect people who may not otherwise be able to establish market-based injuries. A number of states have included statutory damages in publicity legislation with the express purpose of protecting ordinary people. The NO FAKES draft contains statutory damages. It would provide for \$5,000 in damages *per violation* (or the traditional monetary damages, if they're greater). Every creation or single distribution of an unauthorized "likeness" could cost the offender \$5,000. Putting these two provisions together—along with the potential for punitive damages, and discretionary

---

<sup>59</sup> See, [Notice of Termination | U.S. Copyright Office \(https://www.copyright.gov/recordation/termination.html\)](https://www.copyright.gov/recordation/termination.html)

<sup>60</sup> See, e.g., Jennifer E. Rothman, What Happened to Brooke Shields was Awful. It Could Have Been Even Worse, SLATE, April 18, 2023, at <https://slate.com/human-interest/2023/04/brooke-shields-hulu-documentary-child-stars.html>.

<sup>61</sup> [California Child Actor's Bill - Wikipedia \(https://en.wikipedia.org/wiki/California\\_Child\\_Actor%27s\\_Bill\)](https://en.wikipedia.org/wiki/California_Child_Actor%27s_Bill)

attorneys' fees the NO FAKES Act would expand the right of publicity and act as a strong disincentive for any such unauthorized digital replication. Whatever the damages amount we believe there should be a mechanism for inflation adjustments as well.

The scope of the draft's liability provision is vast and based on strict liability. As drafted, general "consumer" users of GAI software could be found liable for violating the digital replica right when they use generative AI programs. Further, the proposed bill apparently targets GAI models as well.

### **Individuals Covered**

Whatever is enacted should apply only to individuals who were or are citizens or residents of the United States within the lookback period, whatever that length of time might be, with reciprocity for the citizens or residents of countries that similarly protect protected American citizens or residents.

### **Communications Decency Act Section 230**

Section 230 of the CDA bolsters First Amendment safeguards of online platforms for user generated content services by guaranteeing protection against liabilities arising from their content moderation efforts by making sure they would not be treated as publishers of third-party content. Unlike publications like newspapers that are accountable for the content they print, online services are relieved of this liability for UGC. Section 230 has two purposes: one was to encourage the unfettered and unregulated development of free speech on the Internet, and the other was to allow or encourage online services to implement their own standards for policing content, and providing for child safety. Section 230 empowers UGC services to moderate content and refine moderation methods, including via algorithms, without acquiring the legal awareness or "knowledge" that could expose them to expensive lawsuits and significant liability.<sup>62</sup>

Section 230 does not necessarily mandate a binary "take down" or "leave alone" regime. There are other solutions available to online platforms, such as warnings that content might be fakes.<sup>63</sup>

Designating the digital replica right as "intellectual property" for purposes of the CDA Section 230, would facilitate the removal of infringing content and the ability to obtain damages from online platforms. Of course there are debates about whether Section 230 should be amended to, among other things, have claims of fakes and rights of publicity, require online platforms to take such claimed content down, similar to how the DMCA operates with regard to claims of copyright infringement. There should be meaningful consideration given to this provision of the proposed NO FAKES act or any law involving digital replication rights and prohibitions.

---

<sup>62</sup> See, e.g., [Section 230 - Wikipedia \(https://en.wikipedia.org/wiki/Section\\_2300\)](https://en.wikipedia.org/wiki/Section_2300), and Goldman, Eric, Why Section 230 Is Better Than the First Amendment (November 1, 2019). *Notre Dame Law Review*, Vol. 95, No. 33, 2019, Available at SSRN: <https://ssrn.com/abstract=3351323> or <http://dx.doi.org/10.2139/ssrn.3351323>

<sup>63</sup> See, e.g., Eric Goldman, *Content Moderation Remedies*, 28 MICH. TECH. L. REV. 1 (2021). Available at: <https://repository.law.umich.edu/mtlr/vol28/iss1/2>

There is currently a circuit split about whether state right of publicity claims fall within the immunity provisions of the CDA §230 or instead fall under its exception for intellectual property laws.<sup>64</sup> This matters because if the exception does not apply, it may be difficult to get platforms to take down infringing content, and it can be hard to track down individuals who initially created or circulated works. However, that issue could be addressed more directly by amending §230 to clarify that state right of publicity and privacy-based appropriation claims are not immunized by §230.

### **First Amendment Concerns**

The current state of the First Amendment's application to right of publicity claims is uncertain and differs substantially from jurisdiction to jurisdiction with at least five different approaches currently employed, according to Prof. Rothman.<sup>65</sup> The Electronic Frontier Foundation has written "courts have struggled to develop a coherent test for how the First Amendment should apply in these cases. In fact, some courts have abandoned traditional tests that helped make sure that publicity rights claims couldn't be used to shut down protected speech that happened to refer to a celebrity or use her likeness in reasonable ways (in biopics, for example)."<sup>66</sup> There are concerns that the NO FAKES Act (or any law involving digital replication rights and prohibitions) could impact freedom of expression and limit creative expression. For example, it could potentially restrict the use of digital tools to create artistic works involving the likeness of individuals. It does, properly, include exceptions for works protected by the First Amendment, such as sports broadcasts, documentaries, biographical works, or for purposes of comment, criticism, or parody. But the boundaries of these exceptions could be subject to interpretation and legal dispute. Moreover, the Act could potentially chill political speech and suppress creative expression if speakers are forced to guess what crosses the line into regulated "manipulated" media. It's important to note that these are potential impacts, and the actual effects would depend on how the any law involving digital replication rights and prohibitions is interpreted and enforced if it becomes law. It's a matter of balancing between protecting individual rights and fostering innovation and creativity.<sup>67</sup>

---

<sup>64</sup> Compare, e.g., *Hepp v. Facebook*, 14 F.4th 204 (3d Cir. 2021) (<https://casetext.com/case/hepp-v-facebook-2>) (holding that Section 230 did not bar a state right of publicity claim from proceeding against Facebook) with *Perfect 10, Inc. v. CCBill LLC*, 488 F.3d 1102 (9th Cir. 2007) ([https://scholar.google.com/scholar\\_case?case=4735249074019268133](https://scholar.google.com/scholar_case?case=4735249074019268133)) (holding that a right of publicity claim could not proceed because of Section 230 immunity since the claim was under state rather than federal law.

<sup>65</sup> [Prof-Rothman-Comments-to-Copyright-Office-on-Right-of-Publicity-and-AI-October-2023.pdf](https://rightofpublicityroadmap.com/wp-content/uploads/2023/10/Prof-Rothman-Comments-to-Copyright-Office-on-Right-of-Publicity-and-AI-October-2023.pdf) ([rightofpublicityroadmap.com](https://rightofpublicityroadmap.com)) <https://www.regulations.gov/comment/COLC-2023-0006-8229> and at <https://www.regulations.gov/comment/COLC-2023-0006-8229>

<sup>66</sup> [A Broad Federal Publicity Right Is a Risky Answer to Generative AI Problems | Electronic Frontier Foundation \(eff.org\)](https://www.eff.org/deeplinks/2023/07/broad-federal-publicity-right-risky-answer-generative-ai-problems) (<https://www.eff.org/deeplinks/2023/07/broad-federal-publicity-right-risky-answer-generative-ai-problems>)

<sup>67</sup> With regard to ROP and the First Amendment, see also: [The Right of Publicity vs. the First Amendment: A Property and Liability Rule Analysis \(indiana.edu\)](https://www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=1637&context=ilj) (<https://www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=1637&context=ilj>); [The First Amendment and the Right\(s\) of Publicity by Robert Post, Jennifer E. Rothman :: SSRN](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3553946) ([https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3553946](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3553946)); [The Right of Publicity: Privacy Reimagined for a Public World \(Introduction\) by Jennifer E. Rothman :: SSRN](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3188507) ([https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3188507](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3188507))

**Style-Based Claims and 17 U.S.C. § 114(b)**

Professor Rothman writes “The First Amendment provides latitude for works in the style of or same genre as that of others. So, there should not be liability for works merely because they are in the “style” of a Taylor Swift or Drake recording when there is no confusion as to the use of the performers’ voice or participation, and there is no false advertising using their names or likenesses.” We agree on the premise provided that there is no confusion or false advertising. Professor Rothman continues “This is largely the state of the law today, where liability based on sound-recordings under right of publicity laws stems from using confusingly similar voices, not the mere use of a similar vocal or musical style.”<sup>68</sup> <sup>69</sup>

There are legitimate concerns, of course, when vocal style is basically identical to that of an individual. But what to do about tribute and cover bands? According to the RIAA, 2023 saw an eruption of unauthorized GAI vocal clone services that “infringe not only the rights of the artists whose voices are being cloned but also the rights of those that own the sound recordings in each underlying musical track. This has led to an explosion of unauthorized derivative works of our members’ sound recordings which harm sound recording artists and copyright owners. Several of these services are located outside of the United States, including Voicify.ai with a web ranking of 22,033, with 8.8 million visits in the past year This site markets itself as the “#1 platform for making high quality AI covers in seconds!” and includes AI vocal models of sound recording artists, including Michael Jackson, Justin Bieber, Ariana Grande, Taylor Swift, Elvis Presley, Bruno Mars, Eminem, Harry Styles, Adele, Ed Sheeran, and others, as well as political figures including Donald Trump, Joe Biden, and Barak Obama. The service apparently stream-rips the YouTube video selected by the user, copies the acapella from the track, modifies the acapella using the AI vocal model, and then provides to the user unauthorized copies of the modified acapella stem, the underlying instrumental bed, and the modified remixed recording.” The RIAA asserts this unauthorized activity infringes copyright as well as infringing the sound recording artist’s rights of publicity.<sup>70</sup>

A risk in enacting such laws and regulations that are not well considered and well drafted is that of outlawing the creation of work in a known style, where that subsequent work is not a copy of the sui generis work, or in US copyright terms is not substantially similar or otherwise infringing of a work. If we did that whoever painted anything other than the first impressionistic painting, cubistic painting, or every Madonna and Child painting, or whoever created a sound recording, or performed or distributed music with or in the style of or using elements of musical styles or genres known as trap or reggaeton, for example, might be sued.

---

<sup>68</sup> See, e.g., *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093 (9th Cir. 1992) (allowing Lanham Act false endorsement claims and state publicity claims when an advertisement used a singer who copied the vocal sound of Tom Waits in ways that confused listeners as to his participation or sponsorship); *Midler v. Ford Motor Co.*, 849 F.3d 460 (9th Cir. 1988) (allowing liability under right of publicity law for a sound-alike performance when listeners thought Bette Midler was singing and “the distinctive voice of a professional singer is widely known and is deliberately imitated in order to sell a product.”). Although these cases are sometimes understood as protecting mere style, both had evidence of confusion as to the performers actually singing and involved intentional evocation of the singers’ identities for use in commercial advertising for products. When this is not the case, it is appropriate to give latitude for similar styles and recordings of licensed musical compositions, otherwise the first person to record a song will get a monopoly in the work in contravention of copyright law and free speech.

<sup>69</sup> <https://www.regulations.gov/comment/COLC-2023-0006-8229>

<sup>70</sup> <https://www.regulations.gov/comment/USTR-2023-0009-0019>

We don't want someone who videotapes and distributes on YouTube a video of them doing the moonwalk, and monetizing that video, for example, to be sued by the Michael Jackson estate for violating the style portion of a NILVS right.<sup>71</sup>

That said, our position is that performers' voices and likenesses must only be used with their consent and fair market compensation for specific uses.

### **GAI Generated Works**

Some Digital Service Providers (DSPs) are uploading and making available GAI-generated content that does not infringe artists' likeness or other rights. Just because those works might displace human created works does not mean those GAI works should not be allowed, or that human created works should receive income or royalties in association with such works. It should be up to the marketplace, e.g., consumers of entertainment and information, as to whether they wish to consume GAI works, human created works, or a combination. That said, we do believe the Artist Centric approach to licensing DSP's as pioneered by Universal Music Group in their deal with Deezer<sup>72</sup> is a good (first) step in enabling GAI music and human authored music to live side by side. The Artist Centric approach values engagement that is initiated by the consumer at a multiple of passive engagement such as when music is served to the consumer algorithmically. This is intended as compensation for the brand value of an artist to the consumer where the brand signals quality of a certain type to a certain audience. This approach can similarly be applied to GAI where human artistry receives a multiple of value of unbranded GAI content. It is, of course, a venn diagram where GAI content becomes famous and popular.

### **Mitigation of Bias**

Certain music genres, languages, and folk and cultural artifacts and history of various countries and regions may be under-represented or not represented at all in the training of GAI tools, and this should be avoided. The tension in this position arises if (a) licensing is negotiated rather than blanket / compulsory, and/or (b) rights holders of certain genres of music, for example, opt out of allowing their works to be used in training. For example, what if the rights holders of most surf music,<sup>73</sup> or the songs and music of Tibet,<sup>74</sup> Bulgaria,<sup>75</sup> or of Klezmer recordings, refused to allow their sound recordings to be used in training GAI tool developers should strive to mitigate against bias during the data training process, ensuring that diverse and representative data is used.

### **Provenance<sup>76</sup>**

---

<sup>71</sup> Of course there have been lawsuits regarding copyright infringement of choreography and Yoga. That's a related, but different cause of action.

<sup>72</sup> <https://www.musicbusinessworldwide.com/popular-artists-to-get-royalty-boost-on-deezer-as-platform-strikes-artist-centric-agreement-with-universal-music-group/>

<sup>73</sup> [https://en.wikipedia.org/wiki/List\\_of\\_surf\\_musicians](https://en.wikipedia.org/wiki/List_of_surf_musicians)

<sup>74</sup> <https://folkways.si.edu/songs-and-music-of-tibet/world/album/smithsonian>

<sup>75</sup> [добре дошли Song of the Crooked Dance - Early Bulgarian 78 rpm recording industry - history https://songofthecrookeddance.com/history/history.htm.](https://songofthecrookeddance.com/history/history.htm)

<sup>76</sup> Based in part on point 6 of Human Artistry Campaign <https://www.humanartistrycampaign.com/>

There should be accurate recordkeeping of copyrighted works, performances, and likenesses, including the way in which they were used to develop and train any GAI model or tools. Algorithmic transparency and identification of a work’s provenance in ingestion and outputs are important to AI trustworthiness. Stakeholders should work collaboratively to develop standards for technologies that identify the input used to create GAI-generated output.

Content generated largely by GAI should be so labeled,<sup>77</sup> similar to truth in advertising laws, and should describe prompts and inputs based upon the names of authors (e.g., “create a symphonic sound recording about Albert Einstein composed in the style of Phillip Glass, performed in the style of the New York Philharmonic with Yo-Yo Ma playing the cello, and conducted by Gustavo Dudamel) and methodology used to create it -- informing consumer choices, and protecting authors (similar to moral rights) and rights holders. (Note, except in the instance of names, such as above, where trade secrets might be used in creating, modifying, refining, and selecting outputs, and copyright registration is not being applied for, such trade secrets might be protected from disclosure.) There should also be “watermarking” of GAI generated content, such as by Adobe and other members of The Coalition for Content Provenance and Authenticity ([C2PA](#)), an Adobe-backed consortium, and Google where a technology, called SynthID<sup>78</sup> embeds a watermark directly into images created by Imagen, one of Google’s latest text-to-image generators. The GAI-generated label remains regardless of modifications like added filters or altered colors.<sup>79</sup> Another approach would be to use an open-source industry supported blockchain solution where recordings have a mathematical hash recorded describing their initial or native state at creation (such as commercial music) or acquisition (such as images from a camera). This initial mathematical hash could theoretically then be compared to subsequent mathematical hashes in downstream works to determine how much of the original was used in the output.

Although we use the term generated “largely by GAI,” identifying the amount/percentage of use of a certain work in a given output may not be possible, as the algorithms learn the patterns of the work, but do not use the work itself to create the new one.

### 3. POTENTIAL SOLUTIONS: CONTROL, COMPENSATION, AND CREDIT

While GDMG believes the use of copyrighted works for GAI training invokes an exclusive right (except as may be covered by fair use, fair dealing, and text and data mining exceptions), and authorization may be required from rights holders in certain circumstances with regard to outputs, because of the considerations of fair use, not wishing to further empower incumbents, the data-sets being huge, and there not being adequate Copyright Management Information, we recommend the industry contemplate several different business models that in essence are tantamount to providing degrees of societally desirable control, compensation, and credit for human authors and rights-holders. We suggest a two pronged

---

<sup>77</sup> We suggest a prominent disclosure along the lines of “Some [copy][marketing materials][illustrations][photographs][voice-overs][videos][editorial content][etc.] authored by people/humans (or entities) with some partial assistance by generative artificial tools with human authorship and creative contributions and control,” if that is accurate.

<sup>78</sup> <https://www.deepmind.com/blog/identifying-ai-generated-images-with-synthid>

<sup>79</sup> <https://www.cnn.com/2023/08/30/tech/google-ai-images-watermark/index.html>

approached to licensing.

## 1. Inputs

- a. Regardless of whether or not the ingestion of copyrighted works by GAI tools is fair use, GAI companies should pay a percentage of their income to rights holders on a compulsory blanket license basis for the rights to train on copyrighted content and human NILVS. This is similar in concept to performing rights organizations for music compositions. Without the copyrighted content, assuming no fair use, fair dealing, or TDM exceptions, the commercial GAI tools would have been trained only on works in the public domain or works that are licensed under a Creative Commons type license, allowing for commercial uses potentially leading to bias due to a smaller training set. This income could be split among rights-holders based on market share.

## 2. Prompts & Outputs

- a. Downstream GAI versions that provenance and the Copyright Office policy deem a derivative work or to include Digital Replica / NILVS rights would need to be directly licensed. We suggest a model similar to how TikTok tracks creations of videos and streams of those videos as two separate metrics and weights in the revenue sharing calculation. These outputs could be further paid for in proportion of what is used, with the value of the use in proportion to the totality of the output. This is similar to the YouTube vertical and horizontal splitting of economics. A share is provided to each type of IP in the output and among each type of IP the share is split pro-rata if there is more than one owner of the same type of IP.
  - i. This might include modifiers to the weighting of revenues to certain works based on the specificity of the prompt engineering. This is similar to the ‘Artist Centric’ model that Universal Music Group has pioneered with Deezer providing for a multiplier providing a heavier weighting to streams occurring on content where the user searched for the artist or song directly as opposed to the music being algorithmically served to the user by the service. In the GAI context an output made from a specific prompt using an artist’s or works name might be provided a heavier weighting when the compulsory license income is being divided up among participants.
  - ii. Of course, to the extent that an entity wishes to voluntarily negotiate licenses for ingestion or training, whether as a hedge against lawsuits or otherwise, that would be up to them.<sup>80</sup>
- b. Notwithstanding the foregoing, GAI tools that generate **deep-fakes** of works should not be, and are not permissible without the consent(s) of the rights holder(s), whether under current copyright law where, for example, the output

---

<sup>80</sup> See, e.g., Adobe Firefly. [https://www.adobe.com/be\\_en/sensei/generative-ai/firefly/enterprise.html](https://www.adobe.com/be_en/sensei/generative-ai/firefly/enterprise.html) Firefly for Enterprise - Adobe. "Enterprises also have the opportunity to obtain an IP indemnity from Adobe for content generated by select workflows powered by Firefly.\* \* We are looking into providing IP indemnification only for text-to-image features in Adobe Express and Adobe Stock under certain Adobe offers. Terms will apply."

substantially resembles the expression of the style of a rights holder, or rights of publicity, privacy, or against impersonation.<sup>81</sup>

---

<sup>i</sup> Jason founded GDMG and is its Chairman. Matt was temporary co-general counsel and is special counsel to GDMG. He has been GC for a variety of companies, including Morpheus of MGM v. Grokster, has taught IP Licensing at Pepperdine SOL, and has practiced law for many years. <https://www.linkedin.com/in/mattneco>

---

<sup>81</sup> See, e.g., WAVs.AI, a site where one may access deepfake music with “artists” like Bad Bunnai, Draik, Ariana Granday, Freddie Mercurai, etc. and Discord server ‘AI Hub’. See also, the section titled **“Impersonations/Right of Publicity and Privacy.”**