

Submitted by Susan Chertkof on behalf of American Association of Independent Music (“A2IM”) and Recording Industry Association of America, Inc. (“RIAA”)

**Before the
UNITED STATES COPYRIGHT OFFICE
Washington, D.C.**

In the Matter of:

Artificial Intelligence and Copyright

Docket No. 2023–6

**COMMENTS OF THE
AMERICAN ASSOCIATION OF INDEPENDENT MUSIC AND
RECORDING INDUSTRY ASSOCIATION OF AMERICA, INC.**

The American Association of Independent Music (“A2IM”) and the Recording Industry Association of America, Inc. (“RIAA”) are pleased to provide these comments in response to the Notice of Inquiry (“NOI”) published by the Copyright Office (the “Office”) on August 30, 2023. *See* 88 Fed. Reg. 59942.

The American Association of Independent Music is a 501(c)(6) not-for-profit trade organization headquartered in New York City that exists to support and strengthen the independent recorded music sector and the value of recorded music copyrights. Membership currently includes a broad coalition of hundreds of independently owned American music labels. A2IM represents these independently owned small and medium-sized enterprises' interests in the marketplace, in the media, on Capitol Hill, and as part of the global music community. In doing so, it supports a key segment of America's creative class that represents America's diverse musical and cultural heritage. Billboard Magazine identified the independent music label sector as over 40 percent of the music industry’s global recorded music revenue in 2020 based on copyright ownership.

The Recording Industry Association of America is the trade organization that supports and promotes the creative and commercial vitality of music labels in the United States, the most vibrant recorded music community in the world. RIAA’s membership – which includes several hundred companies, ranging from small-to-medium-sized enterprises to global businesses – creates, manufactures, and/or distributes sound recordings representing the majority of all lawfully recorded music consumption in the United States. In support of its mission, the RIAA works to protect the intellectual property and First Amendment rights of artists and music labels; conducts consumer, industry, and technical research; and monitors and reviews state and federal laws, regulations, and policies.

Introduction

Together A2IM and RIAA represent a wide swath of music creators and copyright owners who collectively create the vibrant soundtrack that accompanies our lives and is listened to by people across the globe. Human creative expression is at the core of what our members do and support, and it is vital for our nation’s culture and economy.

The music community contributes significantly to the U.S. economy and to U.S. jobs. In 2021, the value added to the GDP by the total copyright industries, of which we are a vital part, exceeded \$2.9 trillion, accounting for 12.52% of the U.S. economy.¹ In addition, the total copyright industries employed nearly 16.1 million workers in 2021, accounting for 8.14% of all U.S. employment.² The music industry itself creates jobs and boosts the economy in all 50 states.³

As with other new technologies, the music community lives on the forefront of, and is building and inspiring, evolutions in artificial intelligence (“AI”) technology. AI already is playing a role as a tool to assist the creative process and will increasingly do so, allowing for a wider range of people to express themselves creatively. Aside from the potential artistic impact, AI is also poised to make the lives of artists easier. Touring with greater efficiency, finding new fans, streamlining licensing, monetizing music in new ways never imagined – AI can do great things for artists. We embrace these technological advances in support of creativity. However, we have already experienced harm from the unethical development and deployment of AI systems that unfairly exploit our artists’ and our members’ expression, creative contributions, names, images, voices, and likenesses without their consent and without compensation.⁴

As signatories to the Human Artistry Campaign (“HAC”),⁵ we call upon policy makers, AI developers, and those that deploy AI to take into account the following principles:

- (i) technology has long empowered human expression, and AI will be no different;
- (ii) human created works will continue to play an essential role in our lives;
- (iii) use of copyrighted works and the use of voices and likenesses of professional performers requires authorization and free-market licensing from all rightsholders;
- (iv) governments should not create new copyright or other IP exemptions that allow AI developers to exploit creations without permission or compensation;
- (v) copyright should only protect the unique value of human intellectual creativity;
- (vi) trustworthiness and transparency are essential to the success of AI and protection of creators; and
- (vii) creators’ interests must be represented in policy making.⁶

¹ Robert Stoner et al., *IIPA, Copyright Industries in the U.S. Economy, 2022 Report*, Secretariat Economists, prepared for the International Intellectual Property Alliance at 8, Dec. 2022, available at: https://www.iipa.org/files/uploads/2022/12/IIPA-Report-2022_Interactive_12-12-2022-1.pdf.

² *Id.*

³ 50 States of Music, <https://50statesofmusic.com/>.

⁴ For example, several songs featuring AI-cloned vocals have gone viral, infringing the rights of the artists whose voices are being cloned and the rights of those that own the musical composition and the sound recording in each underlying track. *See, e.g.*, Dani Di Placido, *Thanks to AI, Fake Kanye and Drake Songs are Going Viral on TikTok*, Forbes (Apr. 24, 2023, 10:07 AM), <https://www.forbes.com/sites/danidiplacido/2023/04/24/ai-generated-songs-that-sound-like-kanye-and-drake-are-going-viral-on-tiktok/?sh=1f9bfcf13531>. In addition, it bears observing that many of the leading generative AI companies are either recipients of enormous investments from dominant internet platform companies or themselves affiliated with such companies, raising the risk that the same competitive harms visible today in online search, social media, and user-generated content platforms will repeat themselves as AI services are deployed.

⁵ *See* Human Artistry Campaign, <https://www.humanartistrycampaign.com/> (last visited on Oct. 27, 2023).

⁶ *Id.*

As key stakeholders in AI policy development, we welcome the opportunity to share our views on AI and copyright with the Office.

Responses to Questions Asked in the NOI

- 1. As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?**

The recorded music industry is, and has always been, a tech-forward business, from the invention of the phonograph, through the eras of vinyl, tapes, CDs, and now streaming, social media, and user-generated content (“UGC”). Record companies appreciate the valuable role that new technologies, including AI, can play in the creative process. In fact, AI and machine learning are already in use in many facets of music production and distribution. For example, Apple Logic Pro X can be used to generate drum tracks and Captain Plugins can be used to generate chord progressions.

AI can be enormously beneficial when it is implemented in a responsible, respectful, and ethical manner. Like every new technology, AI will undoubtedly push creative boundaries and help shape recording artists’ visions and expand their commercial reach.⁷ We embrace AI’s potential as a tool to *support* human creativity, provided that it is not used to *supplant* human creativity.

As exciting as AI is, by and large, we are not seeing it implemented in a responsible, respectful, and ethical manner. In particular, the unauthorized ingestion of our members’ copyrighted works for purposes of training generative AI systems amounts to copyright infringement on a massive scale and causes significant economic harm to our members and their sound recording artists. When AI is designed and/or used to appropriate an artist’s name, image, voice, or likeness – without authorization – it is equally problematic. An artist’s work, persona, and “brand” are deeply personal and reflect years of significant investments of time, money, and effort by the artist and their label support team.

Another risk of AI technology is that machine-generated material can be produced at a speed and scale that creates the very real potential for that material to overrun the marketplace, crowd out

⁷ For example, a foreign artist signed to one of our member companies used a generative AI system to train on recordings of his vocals – allowing him to simultaneously release his single in six languages – in his own voice – on the same day. In that example, the ethically trained tool enhanced and extended the artist’s creative intent – with his consent – enabling him to reach new markets and fans. *See also* Jeff Benjamin, *HYBE’s ‘New’ K-Pop Artist MIDNATT Is Using AI Technology for a Remarkably Human Purpose*, Billboard (May 17, 2023), <https://www.billboard.com/music/pop/artificial-technology-kpop-artist-midnatt-hybe-interview-1235329459/>; Ashley King, *The Beatles’ ‘Last Song’ Is Being Released With the Help of AI – McCartney Calls It ‘A Genuine Beatles Recording’*, Digital Music News (Oct. 27, 2023), <https://www.digitalmusicnews.com/2023/10/27/the-beatles-last-song-to-release-with-the-help-of-ai/>.

human-created work, and generally devalue works created by human beings.⁸ These dynamics, which will have potentially wide-ranging effects on the music streaming market, are already starting to be felt. Our members have made it known that they are concerned about the proliferation of low-quality, machine-generated audio that is available on digital services, making it harder for fans to find the artists they already love and discover new artists that will ripen into future favorites. Today, “music” that is purely machine-made has little appeal to genuine fans but can divert the flow of royalties and engagement away from human creators.⁹ The proliferation of industrialized machine-made audio also devalues human artistry.

This is precisely why the Constitution includes copyright protection¹⁰ -- to “incentiviz[e] individuals to create and invent. The act of human creation—and how to best encourage human individuals to engage in that creation, and thereby promote science and the useful arts—was thus central to American copyright from its very inception. Non-human actors need no incentivization with the promise of exclusive rights under United States law, and copyright was therefore not designed to reach them.”¹¹

As you develop recommendations for Congress on the rules of the road for AI, we urge the Office to ensure that sound recording artists and the businesses that support them are incentivized to continue producing original, creative works that break new ground, uplift human creative expression and enrich our world.

2. Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?

As described more fully below, our industry faces a number of unique challenges, some legal and some technological. We also have a positive track record of voluntary marketplace licensing deals in the context of new technologies – including full catalog licenses with streaming music services and user-generated content, social media, fitness services, as well as services that offer other innovative business models – that positions us well to negotiate voluntary marketplace deals with AI developers.

Section 114(b) of the Copyright Act

Legally speaking, sound recordings are unique among copyrighted works because Section 114(b) of the Copyright Act limits the scope of protection for sound recordings to uses “that directly or

⁸ See, e.g., Daniel Tencer, *AI Music App Boomy Has Created 14.4M Tracks to Date. Spotify Just Deleted a Bunch of Its Uploads After Detecting ‘Stream Manipulation,’* Music Business Worldwide (May 3, 2023), <https://www.musicbusinessworldwide.com/ai-music-app-boomy-spotify-stream-manipulation/> (“According to Boomy’s website, since the AI startup was founded in the U.S. in 2019, its users have created a whopping 14.4 million songs, which, the firm boasts, accounts for “around 13.78% of the world’s recorded music.”).

⁹ AI-generated music can also be used as a tool for fraud and to illegally siphon royalties away from artists and rightsholders. As was recently reported, we are also starting to see uploads of AI-generated tracks followed by bots that are used to create “fake listens” of those tracks through the practice of so-called “stream manipulation.” Our members are already working in partnership with platforms and distributors to combat the problem of stream manipulation across the board, and this issue is very much on our radar.

¹⁰ U.S. Const. art. I, § 8, cl. 8.

¹¹ *Thaler v. Perlmutter*, No. 22-1564 (BAH), 2023 WL 5333236, *4 (D.D.C. Aug. 18, 2023).

indirectly recapture the actual sounds fixed in the recording.” However, to the extent that Section 114(b) may be relevant to the Office’s analysis, the Office should recognize that Section 114(b) does not permit the reprocessing of a copyrighted recording through artificial intelligence systems without liability for copyright infringement.

The idea for Section 114(b) dates back to the earliest efforts to provide copyright protection for sound recordings.¹² Section 114(b) and its predecessors were designed to protect copyright owners from the copying of their recorded performances by technological processes of any kind, while merely providing an exception from the normal scope of copyright protection to allow others to independently render and record different performances. During the general revision of the Copyright Act, the Register explained that the draft bill at that time would protect recordings against reproduction “by any method” while it would “not offer any protection against a separate recording of another performance in which those sounds are imitated.”¹³ When the general revision process dragged on, and sound recording protection was accelerated, Congress again recognized this distinction, granting exclusive rights to reproduce recordings in a way “that directly or indirectly recaptures the actual sounds fixed in the recording,” but not as to “duplication of another sound recording that is an independent fixation of other sounds.”¹⁴ In doing so, Congress noted that “[a]ny unauthorized manufacturer who wishes to produce a record containing the same songs may do so by . . . making the same investment in production and talent as is being done by the authorized record companies.”¹⁵

These principles were codified in current Section 114(b). Today, the sound recording reproduction right is limited to “copies that directly or indirectly recapture the actual sounds fixed in the recording,” but that is distinguished from “duplication of another sound recording that consists entirely of an independent fixation of other sounds.”¹⁶ In enacting this language, Congress explained its intent to provide protection against reproduction of copyrighted recordings “by repressing, transcribing, recapturing off the air, or any other method,” while “not prevent[ing] a separate recording of another performance in which [the original] sounds are imitated . . . even where one performer deliberately sets out to simulate another’s performance as exactly as possible.”¹⁷

In the intervening half century, the technology of sound recording reproduction has evolved from straightforward analog duplication to include digital mixing, sampling, and encoding techniques.

¹² See Subcomm. on Patents, Trademarks, and Copyrights, H. Comm. on the Judiciary, 86th Cong., Study No. 26, *The Unauthorized Duplication of Sound Recordings* 37, 37 n.354 (Comm. Print 1961 by Barbara A. Ringer) (“Throughout the [1936 and 1947] hearings there was a great deal of confusion between protection against the actual reproduction of a particular recording and protection against imitation or mimicry of a general style or manner of performance.”).

¹³ H. Comm. on the Judiciary, 89th Cong., *Supplementary Register’s Report on the General Revision of the U.S. Copyright Law* 52 (Comm. Print 1965) (emphasis added).

¹⁴ Pub. L. No. 92-140, 85 Stat. 391 (Oct. 15, 1971) (emphasis added).

¹⁵ S. Rep. No. 92-72, at 6 (1971).

¹⁶ 17 U.S.C. § 114(b) (emphasis added).

¹⁷ H.R. Rep. No. 94-1476, at 106 (1976).

As the technology has changed, courts have rightfully recognized that digital processing of recorded sounds is simply another way of using the actual sounds involved.¹⁸

Ingestion of copyrighted sound recordings to develop an AI model that generates derivative audio outputs is simply the latest new technological method to process recordings. Neither the model nor its outputs “consists entirely of an independent fixation of other sounds.”¹⁹ Clearly, there is no “separate recording of another performance.”²⁰ And nobody has made “the same investment in production and talent” as the copyright owner.²¹ Rather, the model and its outputs depend on the copyrighted inputs, which they “directly or indirectly recapture.”²² While the new technology of generative AI may process the sounds of copyrighted recordings in a more sophisticated and obscure manner than the analog duplication techniques available a half century ago, they are essentially remixing the copyright owner’s sounds, rather than making the kind of new recording of different sounds that Section 114(b) was intended to enable.

Voice-Cloning

Another issue that is somewhat, but not entirely, unique to our industry is voice-cloning, which is typically used to create unauthorized recordings that appear to feature one or more popular recording artists who did not actually perform on those recordings. Artists need the right to control the use of their voices, protect the integrity of their artistic visions and images, and preserve the value of their brands in the marketplace. And third parties should not be permitted to steal an artist’s voice without that artist’s permission.

When AI is used to simulate a particular recording artist’s voice without that artist’s consent, a host of laws protecting names, images, and likenesses are implicated, including Section 43(a) of the (federal) Lanham Act and state rights of publicity and privacy. While these laws are not the direct focus of the Office, it is impossible to discuss the practice of AI “voice cloning” without at least mentioning them. Proper interpretation and enforcement of these federal and state laws is essential.

Unauthorized voice-cloning also implicates a variety of copyright violations. For example, voice-cloning AI models are typically built on datasets consisting of vocal stems that are extracted from copyrighted sound recordings and ingested – whether in whole or in smaller segments – to “fine-tune” the model to improve its voice-mimicking capabilities. Those stems are themselves subject to copyright, and their unauthorized use in this manner – which includes extracting the stems from their source and ingesting the stems to train the AI model – is also infringing. Representations of the sounds in those vocal stems are captured in the fine-tuned vocal clone model, also potentially giving rise to further copyright violation.

¹⁸ See *UMG Recordings, Inc. v. MP3.Com, Inc.*, 92 F. Supp. 2d 349, 350 n.1 (S.D.N.Y. 2000) (rejecting claim “that the simulated sounds on MP3-based music files are not physically identical to the sounds on the original CD recordings”); see also *Capitol Records, LLC v. BlueBeat, Inc.*, 765 F. Supp. 2d 1198, 1204 (C.D. Cal. 2010) (rejecting claims with respect to “psychoacoustic simulations” of copyrighted recordings).

¹⁹ 17 U.S.C. § 114(b).

²⁰ H.R. Rep. No. 94-1476, at 106.

²¹ S. Rep. No. 92-72, at 6 (1971).

²² 17 U.S.C. § 114(b).

Similarly, many voice clones are used to superimpose an artist’s voice over an existing, copyrighted sound recording that the artist never actually performed on themselves. Unauthorized reproduction, distribution, and performance of those recordings obviously infringes the copyrights in the underlying sound recording, and we have been successful in having those recordings taken down from online platforms on that basis.

Violations of Sections 1201 and 1202 of the Copyright Act

Other “unique” issues that affect sound recordings in this context include violations of Section 1201 when stream ripping technology is used to obtain sound recordings for use as datasets and violations of Section 1202 when copyright management information is stripped from the sound recordings in the training/ingestion process (see discussion in response to Question 26); and possibly other violations as well.

Track Record of Innovative, Voluntary Marketplace Deals

Our industry is also unique in one especially important way. Our members have many years of experience negotiating innovative licensing deals that offer digital access to their recordings to global, regional, and domestic music services – including streaming music services (e.g., Spotify, Apple Music), services that offer user-generated content (“UGC”) (including short-form videos) and social media (e.g., YouTube, TikTok (including its predecessor Musical.ly), Instagram, Snap), and digital fitness services (e.g., Peloton, Apple Fitness+) (together “Digital Music Services”) – on a voluntary basis via the commercial marketplace.²³ Voluntary licensing has been the key to the resurgence of U.S. recorded music revenues – which declined by more than half due to rampant piracy at the start of the digital era²⁴ – but which are once again robust enough to fund significant investments by our member companies in their artists and in finding innovative new ways to connect their artists to their fans. Indeed, our members now derive the vast majority of their revenues from services that have been licensed in the free market.²⁵

The wide variety of use cases and business models reflected in our members’ existing licensees demonstrates the industry’s willingness and ability to be innovative and flexible in building new models and addressing new formats. The infrastructure for voluntary market-based licensing

²³ And, of course, they regularly negotiate licenses that allow their sound recordings to be sampled in other recordings along with sync licenses that permit their artists’ sound recordings to be used in television shows, motion pictures and advertising. For a list of the many authorized digital music models and services in today’s marketplace, see *Find Your Music*, Why Music Matters, <https://whymusicmatters.com/>.

²⁴ The industry market value went from \$14.6 billion in 1999 to just \$7 billion a decade later (not adjusted for inflation).

²⁵ Recorded music revenues reached an all-time first-half high of \$8.4 billion in 2023, growing 9.3% at estimated retail value. Digital Music Services continued to flourish at \$7 billion and accounted for 84% of the first-half total, with paid subscriptions still the strongest driver of revenue growth, increasing by more than \$550 million and averaging 96 million subscriptions during the period. See Joshua P. Friedlander & Matthew Bass, *RIAA Mid-Year 2023 Revenue Report*, Recording Indus. Ass’n of Am. (2023), <https://www.riaa.com/wp-content/uploads/2023/09/RIAA-Mid-Year-2023-Revenue-Report.pdf>.

already exists.²⁶ That, plus the licensing principles, practices, and experience our members have developed over the years leaves them uniquely positioned to develop a voluntary licensing framework, that is global in reach, for the new generative AI ecosystem.

- 3. Please identify any papers or studies that you believe are relevant to this Notice. These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.**

Please see Annex A for a list of some papers relevant to this notice of inquiry.

- 4. Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States? How important a factor is international consistency in this area across borders?**

The development of public policy surrounding AI is in its infancy, presenting the U.S. with an important opportunity to lead the world in maintaining strong respect for copyright and the rights of creators. As the Office and Congress consider guidelines and rules for a responsible and safe AI ecosystem, we note the helpful commitments made by the G7, including the Hiroshima Leaders' Communiqué (paragraph 38),²⁷ the Ministerial Declaration of the Digital and Tech Ministers' Meeting (*see* paragraphs 42, 45, and 47),²⁸ and the G7 Digital and Tech Ministers' statement from September 2023 (*see* paragraph 11).²⁹ We also note that in the recent G20 New Delhi Leaders' Declaration, the leaders of the G20 recognized the importance of respecting intellectual property rights in the building of digital public infrastructure.³⁰ Continuing this trend, on October 30, 2023, the Leaders of the G7 agreed upon the Hiroshima Process

²⁶ For independent record labels, this infrastructure includes distributors like Empire, Red Eye, Symphonic, AWAL, The Orchard, ADA, and Ingrooves, digital rights agents such as Merlin and, of course, licensing by the labels themselves.

²⁷ *See* G7 Hiroshima Leaders' Communiqué, ¶ 38, (May 20, 2023), <https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/20/g7-hiroshima-leaders-communicue/> (“We task relevant ministers to establish the Hiroshima AI process, through a G7 working group, in an inclusive manner and in cooperation with the OECD and GPAI, for discussions on generative AI by the end of this year. These discussions could include topics such as governance, safeguard of intellectual property rights including copyrights, promotion of transparency, response to foreign information manipulation, including disinformation, and responsible utilization of these technologies.”).

²⁸ *See* G7 2023 Hiroshima Summit, *Ministerial Declaration the G7 Digital and Tech Ministers' Meeting 30 April 2023*, ¶ 42, 47 (Apr. 30, 2023), http://www.g7.utoronto.ca/ict/2023-ministerial_declaration_dtmm.pdf. Among other things, the digital and technology ministers of the G7 countries declared that the G7 “reaffirm their commitment to promote human-centric and trustworthy AI based on the OECD AI Principles,” “oppose the misuse and abuse of AI to . . . threaten the enjoyment of human rights,” and “plan to convene further G7 discussions on generative AI which could include topics such as governance, how to safeguard intellectual property rights including copyright, promote transparency, address disinformation, including foreign information manipulation, and how to responsibly utilize these technologies.”

²⁹ *See* G7 Hiroshima AI Process, *G7 Digital and Tech Ministers Statement*, ¶ 11 (Sept. 7, 2023), https://www.politico.eu/wp-content/uploads/2023/09/07/3e39b82d-464d-403a-b6cb-dc0e1bdec642-230906_Ministerial-clean-Draft-Hiroshima-Ministers-Statement68.pdf.

³⁰ *See* G20 New Delhi Leaders' Declaration, ¶ 56 (Sept 9-10, 2023), https://www.g20.org/content/dam/gtwenty/gtwenty_new/document/G20-New-Delhi-Leaders-Declaration.pdf.

International Guiding Principles for Organizations Developing Advanced AI System, which provide that organizations should “implement appropriate data input measures and protections for personal data and intellectual property,” further noting that “[a]ppropriate transparency of training datasets should also be supported and organizations should comply with legal frameworks.”³¹ They also agreed upon the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems, which further encourages such organizations to “implement appropriate safeguards, to respect rights related to privacy and intellectual property, including copyright-protected content.”³²

In terms of legislation, the EU’s AI Act, currently under consideration, includes helpful proposals on government review of generative AI models before release, continued assessment of those models, recordkeeping provisions, transparency and labeling obligations, and more. As one of the bill’s key recitals reads: “...it is appropriate for the Commission and the AI Office to monitor and periodically assess the legislative and governance framework of such models and in particular of generative AI systems based on such models, which raise significant questions related to the generation of content in breach of Union law, copyright rules, and potential misuse.”

On the other hand, there are some policies, including ones that were adopted years ago, before the rise of generative AI, that the U.S. should avoid. For example, Japan introduced in 2009, and then extended in 2018, a text and data mining exception to copyright law that permits the exploitation of a copyrighted work for data analysis, with only limited, ambiguous conditions placed upon such exception.³³ Another example is the policy in Singapore, enacted in 2021, which permits copying or communication for computational data analysis, regardless of whether the use is for commercial or noncommercial purposes.³⁴

We also note that, in June 2022, the United Kingdom considered broad text and data mining policies that it subsequently rejected in recognition of the irrevocable harm such policies would inflict upon the U.K.’s creative industries.³⁵ The U.K. House of Commons, Culture, Media and Sports Committee recommended “that the Government does not pursue plans for a broad text and data mining exemption to copyright” and stated further that “[t]he Government should support the continuance of a strong copyright regime in the UK and be clear that licences are

³¹ See *Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI system*, ¶ 11, Oct. 30, 2023, available via <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system>.

³² See *Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems*, ¶ 11, Oct. 30, 2023, available via <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems>.

³³ Japan’s Copyright Act, Act No. 48 of 1970, Art. 30.4. See also Tatsuhiro Ueno, Text-and-data Mining and Copyright slides for ALAI Cong. in Madrid (Sept. 30, 2021), available at http://www.f.waseda.jp/uenot/index_e.html.

³⁴ Singapore Copyright Act of 2021, Act No. 22 of 2021, Div. 7 § 244. See also Albert Kang, *Coming Up in Singapore: New Copyright Exception for Text and Data Mining*, Bird & Bird (Sept. 19, 2021), <https://www.twobirds.com/en/insights/2021/singapore/coming-up-in-singapore-new-copyright-exception-for-text-and-data-mining/>.

³⁵ See *Connected tech: AI and creative technology*, Eleventh Report of Session 2022-23, House of Commons, Culture, Media and Sport Committee (July 18, 2023), <https://committees.parliament.uk/publications/41145/documents/201678/default/>.

required to use copyrighted content in AI. In line with our [the Committee’s] previous work, this Committee also believes that the Government should act to ensure that creators are well rewarded in the copyright regime.”³⁶

The question also asks about the importance of international consistency. While we see this as a valid goal, we urge the U.S. to approach the topic as a thought leader that encourages other countries to maintain broad support for copyrights and to resist any pressure or temptation to join an international race to the bottom.³⁷ Another relevant consideration is finding a way to prevent AI developers from geo-launders. For example, an AI developer should not be able to import an AI model into the U.S. that ingested copyrighted works without authorization by claiming that their AI development occurred in another jurisdiction where they claim the ingestion of copyrighted works is legal. U.S. laws that respect copyright ownership are not going to make the U.S. less competitive. As the White House has stated, “[t]he important progress [of AI automated systems] must not come at the price of civil rights or democratic values.”³⁸ Respect for human rights and democratic values necessarily includes respecting the property rights of authors and artists, as well as respect for copyright as an engine for economic growth.

5. Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.

As a general matter, we think existing copyright laws are sufficient to address the copyright-related issues that have arisen so far in connection with generative AI, although the transparency concerns highlighted in response to Questions 15-17 (and elsewhere in these comments) often render these laws unavailable to us as a practical matter. We do, however, have three non-copyright legislative proposals that address important aspects of our members’ ability to effectively enforce their copyrights and related rights. In addition, there is one issue that may warrant copyright legislation in the future, depending on technological developments. Our legislative proposals include the following:

- Federal legislation to protect voice and likeness is needed to create a baseline legal and economic framework where individuals are shielded from theft and harm, consumers are protected from deception and misinformation, and AI developers and services are provided with clear business certainty. Currently, there is an inconsistent patchwork of state laws. A federal right protecting voice and likeness is needed to establish a uniform “floor” of fundamental rights to ensure an individual or an entity (including an heir) authorized to act on their behalf can take legal action whenever generative AI is employed to use their image, voice, or likeness without their permission, no matter where they live. Individuals’ integrity, livelihoods, and basic human identities are at stake, and

³⁶ *Id.* at 16.

³⁷ In order to maintain our strong respect for intellectual property, the U.S. should avoid not only the copyright text and data mining exceptions contemplated in Japan and Singapore, but also the opt-out text and data mining policy adopted by the European Union. As noted elsewhere in our comments, such an opt-out policy is inconsistent with the exclusive nature of copyrights under U.S. law.

³⁸ See The White House, Office of Science and Technology Policy, *Blueprint for an AI Bill of Rights* (Oct. 4, 2022), <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>.

we must ensure that they are supported and protected – rather than exploited – by technological advancements.

- Congress should enact a new, non-copyright law requiring AI developers to keep detailed records regarding training inputs and other information, such as user prompts, which are vital to promote accountability, facilitate licensing, and enable enforcement. Any recordkeeping law should also include appropriate disclosure requirements and penalties for non-compliance. For a fuller discussion of possible recordkeeping requirements, see our response to Question 15 and its subparts.
- Congress should consider creating a new administrative subpoena process, loosely modeled after the Section 512(h) subpoena process, whereby a copyright owner or a person authorized to act on the owner’s behalf may, by asserting a subjective good faith belief that one or more of the owner’s copyrighted works have been used by an AI developer without authorization, request the clerk of the United States district court to issue a subpoena to an AI developer for identification of any of the copyrighted works that have been reproduced in the training of an AI model. If the AI developer does not comply with the subpoena, either because the developer has not kept the required records or because the developer chooses not to comply, the copyright owner should then be entitled to an evidentiary presumption – sufficient to support the filing of a copyright infringement lawsuit – that the works identified in the subpoena were, in fact, reproduced. And, of course, any applicable statute of limitations should be tolled until the AI developer’s compliance or non-compliance with these obligations has been established. *See* draft language in Annex B.
- Depending on how the methods by which AI systems ingest copyrighted works evolve in the future, it may become necessary for Congress to consider adding a new right to Section 106 of the Copyright Act that gives a copyright owner the exclusive right to ingest a copyrighted work in an AI model or authorize others to do the same.

Training

If your comment applies only to a specific subset of AI technologies, please make that clear.

6. What kinds of copyright-protected training materials are used to train AI models, and how are those materials collected and curated?

Broadly speaking, anything and everything that is available in digital form via the Internet is being ingested by AI models. More specifically, it appears that models are being trained on some combination of open-source datasets, proprietary datasets, public domain materials, and web crawling. While we are aware that some generative AI companies claim that they rely solely on works authorized for their use³⁹ – either by licensing them or by relying solely on

³⁹ *See, e.g.*, Adobe Firefly, <https://www.adobe.com/sensei/generative-ai/firefly.html#:~:text=Trained%20on%20Adobe%20Stock%20images,dataset%20to%20retrain%20Firefly%20models> (claiming that Adobe Firefly is trained on “Adobe Stock images, Openly licensed content, and public domain

works in the public domain – many other generative AI companies take a less responsible approach. Some AI companies scrape vast numbers of copyrighted works from the Internet and elsewhere to harvest for their AI systems without obtaining permission from or compensating the copyright owners of these works. Often, the sites these companies scrape include pirated works.⁴⁰

When it comes to collection and curation, we have witnessed a disturbing practice of willfully disaggregating the creation of datasets for AI training, often by entities that claim to be non-profit or research-focused, and the actual training of AI models, often by for-profit, commercial entities. For example, OpenAI, which occupies an outsized role in the AI marketplace, initially promoted itself as a “research” project but transitioned to a for-profit company in 2019 and is now reportedly worth many billions of dollars.⁴¹ Unless appropriate policies are put into place to prevent it, such disaggregation can easily result in so-called “data laundering,” whereby the developer of a commercial AI model seeks to avoid copyright infringement liability by claiming that the dataset from which it ingested copyrighted works was built for research purposes and using that as a basis to invoke some form of “fair use.” This sort of copyright and data laundering is particularly rampant in countries that have text and data mining exceptions for noncommercial entities, which the U.S. should avoid replicating here (*see* response to Question 4 above).⁴²

6.1. How or where do developers of AI models acquire the materials or datasets that their models are trained on? To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?

See answer to Question 6 above.

6.2. To what extent are copyrighted works licensed from copyright owners for use as training materials? To your knowledge, what licensing models are currently being offered and used?

content”); *Getty Images Launches Commercially Safe Generative Offering*, Getty Images (Sept. 24, 2023), <https://newsroom.gettyimages.com/en/getty-images/getty-images-launches-commercially-safe-generative-ai-offering> (claiming that their Generative AI by Getty Images product is trained solely from Getty Images’ vast creative library). We note that some artists have questioned whether the Adobe licenses cover use of their works for AI purposes.

⁴⁰ *See* Kevin Schaul, Szu Yu Chen, & Nitasha Tiku, *Inside the Secret List of Websites that Make AI like ChatGPT Sound Smart*, The Washington Post (Apr. 19, 2023), <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>.

⁴¹ *See, e.g.*, Felix Salmon, *How a Silicon Valley nonprofit became worth billions*, Axios (Jan. 10, 2023), <https://www.axios.com/2023/01/10/how-a-silicon-valley-nonprofit-became-worth-billions>.

⁴² Some have suggested that this data laundering is likely to extend to global for-profit companies as well. *See, e.g.*, James Love, *We Need Smart Intellectual Property Laws for Artificial Intelligence*, Scientific American (Aug. 7, 2023), <https://www.scientificamerican.com/article/we-need-smart-intellectual-property-laws-for-artificial-intelligence/>. *See also* Justin Hendry, *Google wants copyright law changed for AI data mining*, InnovationAus.com (Apr. 20, 2023), which notes that in response to an Australian inquiry on its copyright law, Google said that the lack of copyright flexibilities in Australia would result in AI-powered products being created elsewhere, and then imported into Australia.

Outside of the music space, we are aware of some companies, such as Adobe and Getty Images, offering products that use generative AI technologies that are said to rely exclusively on authorized use of works – either by licensing copyrighted works or by relying solely on works in the public domain.⁴³ Our member companies have made clear in public statements⁴⁴ and elsewhere that they are “open for business” and welcome discussions regarding licenses of their catalogs for use by legitimate AI businesses. In fact, they are actively engaged in free-market licensing discussions, deals are beginning to get done, and more will get done in the near term.

As is evident from the deals that our members have publicly announced in this space,⁴⁵ every licensing arrangement will be bespoke and dependent on the particular use-case at issue. However, the companies’ licensing history has proven that free-market licensing of sound recordings in the context of innovative new technologies is readily achievable⁴⁶ and beneficial for both licensors and licensees as well as consumers and other end-users. The necessary licensing structures already exist – they are practical, efficient, and have a track record of success. AI developers can cite no valid reason for infringing copyrighted sound recordings.

6.3. To what extent is non-copyrighted material (such as public domain works) used for AI training? Alternatively, to what extent is training material created or commissioned by developers of AI models?

We understand that some AI developers may claim that they use public domain sound recordings.⁴⁷ However, mostly we have seen that developers of generative AI music models use copyrighted sound recordings to develop their AI models. For example, some developers claim to have used licensed music for the model development.⁴⁸ Another claims to have scraped 1.2 million songs, which commentators note are likely copyrighted music.⁴⁹ Another describes its dataset as “audio files of 1,748 pieces of Pop piano from the Internet,” also suggesting that this is

⁴³ See *supra* note 40.

⁴⁴ See, e.g., Mandy Dalugdug, *Google and Universal Music Group in Talks Over Licensing AI ‘Deepfakes’* (Report) (Aug. 9, 2023), <https://www.musicbusinessworldwide.com/google-and-universal-music-group-in-talks-over-licensing-ai-deepfakes-report/>; Daniel Tencer, *Warner Music Group Boss Robert Kyncl on AI, Why Labels Still Exist, and 2 Other Things we Learned from his Q&A at the Code Conference* (Sept. 29, 2023), <https://www.musicbusinessworldwide.com/warner-music-group-boss-robert-kyncl-on-ai-why-labels-still-exist-and-2-other-things-we-learned-from-his-qa-at-the-code-conference/>; Sarah Perez, *Warner Music CEO Robert Kyncl says AI to impact the music industry within the next year*, TechCrunch (Sept. 27, 2023), <https://techcrunch.com/2023/09/27/warner-music-ceo-robert-kyncl-says-ai-to-impact-the-music-industry-within-the-next-year/> (“There’s a very clear analogy to user-generated content – we have a blueprint for this...”).

⁴⁵ Deals that have been announced in this space by our members companies include a deal between UMG and Google, deals between more than one of our member companies and Endel, a company that uses AI to generate wellness soundscapes, and a newly announced deal between UMG and BandLab.

⁴⁶ See discussion in response to Question 2 above.

⁴⁷ See *Concord Music Group, Inc. et al v. Anthropic PBC*, Complaint at ¶ 7, (3:23-cv-01092) (M.D. Tenn) filed Oct. 18, 2023.

⁴⁸ See, e.g., Jade Copet et al., *Simple and Controllable Music Generation*, arXiv, June 8, 2023, at 5, available at <https://arxiv.org/pdf/2306.05284.pdf>; Giorgio Mariani et al., *Multi-Source Diffusion Models for Simultaneous Music Generation and Separation*, arXiv, Feb. 9, 2023, at 3, 9, available at <https://arxiv.org/pdf/2302.02257.pdf>

⁴⁹ Prafulla Dhariwal et al., *JukeBox: A Generative Model for Music*, arXiv, Apr. 30, 2020, at 5, available at <https://arxiv.org/pdf/2005.00341.pdf>; see also Bijan Stephen, *OpenAI introduces Jukebox, a new AI model that generates genre-specific music*, The Verge (Apr. 30, 2023), <https://www.theverge.com/2020/4/30/21243038/openai-jukebox-model-raw-audio-lyrics-ai-generated-copyright>.

copyrighted music.⁵⁰ We have also observed several datasets for AI vocal simulation models that contain the a capella portion of our members’ sound recordings.

6.4. Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)? Please describe any relevant storage and retention practices.

Although we have a very limited window into this issue due to the lack of transparency that AI developers currently provide, it is our understanding that training materials are sometimes retained for document retention or retraining purposes, and sometimes intentionally destroyed after the AI model has been developed. This is one of many reasons that mandatory recordkeeping and disclosure, discussed more fully below in response to Question 15 and its subparts, is imperative. Some of the reasons that AI developers retain training materials include, for example: fine-tuning of foundational models (by further training a foundational model on a smaller, task-specific dataset), de-duping of models, performing quality control tests on their datasets to filter questionable or biased data, and retraining in the event of a court order to purge certain works. For all practical purposes, the use of unauthorized copies of copyrighted works as training materials for AI models *always* and *necessarily* involves reproductions that are sufficiently persistent to infringe copyright (*see* response to Question 7 below).

7. To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained. The Office is particularly interested in:

The views provided below reflect conversations with people at our member companies (and elsewhere) who have personal knowledge of the process by which AI models are trained.

7.1. How are training materials used and/or reproduced when training an AI model? Please include your understanding of the nature and duration of any reproduction of works that occur during the training process, as well as your views on the extent to which these activities implicate the exclusive rights of copyright owners.

Defenders of using unlicensed copyrighted works for AI development contend that the copies made during ingestion are ephemeral and thus not subject to copyright.⁵¹ That is a purely academic argument. In practice, persistent – and therefore actionable – copies of copyrighted

⁵⁰ See Wen-Yi Hsiao et al., *Compound Word Transformer: Learning to Compose Full-Song Music over Dynamic Directed Hypergraphs*, arXiv, Jan. 7, 2021, at 5, available at <https://arxiv.org/pdf/2101.02402.pdf>. See also Flavio Schneider et al., *Mousai: Text-to-Music Generation with Long-Content Latent Diffusion*, arXiv, Jan. 30, 2023, at 5, available at <https://arxiv.org/pdf/2301.11757.pdf>, which claims to have been trained on “2,500 hours of stereo music . . . spanning multiple genres, artists, instruments and provenience in order to maintain a high diversity dataset.”

⁵¹ If that is the case, why do companies like Google, Microsoft, and Adobe all offer to indemnify users of their generative AI products against liability for copyright infringement arising from the use of their tools? See Emilia David, *Google promises to take the legal heat in users’ AI copyright lawsuits*, The Verge (Oct. 12, 2023), <https://www.theverge.com/2023/10/12/23914998/google-copyright-indemnification-generative-ai>.

material are made throughout the training process: first, in compiling and cleaning the dataset,⁵² and then in the model development and fine-tuning. The development and fine-tuning process is an iterative one, so it is often necessary to keep copies of the dataset on hand throughout each iteration. Unauthorized reproduction of copyrighted works violates copyright owners' exclusive right of reproduction under Section 106(1) of the Copyright Act.

7.2. How are inferences gained from the training process stored or represented within an AI model?

We disagree with the premise of the question, which assumes, first, that computers are capable of making “inferences,” implying that computers can engage in some sort of reasoning or thinking, and, second, that only “inferences” are stored and represented rather than reproductions of the copyrighted works ingested by the AI model. We believe that AI models store representations of all or part of our recordings within their models, even if this is in compressed form. As explained more fully in response to Question 2 above, use of copyrighted sound recordings to develop an AI model that generates derivative audio outputs is simply the latest new technological method to process recordings.

Further, examples abound of generative AI models, such as large language transformer and image diffusion models, producing output the same as or materially similar to the content that such models ingested.⁵³ While, this sort of copying/memorization is readily discernable in the cases of image and text generative AI (including, for example, song lyrics),⁵⁴ it surely extends to generative AI music models. Indeed, the authors of the paper announcing Google's generative AI music model, musicLM, acknowledge “the risk of potential misappropriation of creative content associated to the use-case” and “strongly emphasize the need for more future work in tackling these risks associated to music generation,” noting they have “no plans to release models at this point.”⁵⁵ It appears that Google has since released the model in a limited way, notwithstanding these risks.⁵⁶

7.3. Is it possible for an AI model to “unlearn” inferences it gained from training on a particular piece of training material? If so, is it economically feasible? In addition to retraining a model, are there other ways to “unlearn” inferences from training?

⁵² It is worth noting that datasets are made available on various internet sources (e.g., huggingFace, Google Drive, Kaggle, etc.) for others to use for training or other purposes as well.

⁵³ See Section 1 of Annex A for papers describing this phenomenon. See also *Concord Music Group, Inc. v. Anthropic PBC*, Complaint at ¶ 10, (3:23-cv-01092) (M.D. Tenn) filed Oct. 18, 2023 (“Anthropic’s AI models generate output containing Publishers’ lyrics even when the models are not specifically asked to do so.”).

⁵⁴ *Id.*

⁵⁵ Andrea Angostinelli et al., *MusicLM: Generating Music from Text*, arXiv, Jan. 26, 2023, at 8, available at <https://arxiv.org/pdf/2301.11325.pdf>.

⁵⁶ See Ngozi Chukwu, *Google is set to change how you create and listen to music with its AI music generator*, TechCabal (May 12, 2023), <https://techcabal.com/2023/05/12/google-is-set-to-change-how-you-create-and-listen-to-music-with-its-ai-music-generator/#:~:text=Google%27s%20AI%20music%20generator%2C%20MusicLM,required%20to%20join%20a%20waitlist>.

At the outset, we take issue with the premise of the question. As with Question 7.2, this question assumes that computers are capable of “learning” and “unlearning,” an idea we dispute. In addition, the question seems to assume that AI developers should not be required to use “unlearning” tools if doing so would create some sort of economic hardship for them. It also does not ask about the economic hardships creators and copyright owners face when AI developers engage in copyright infringement on a massive scale without any compensation to those upon whose copyrighted works the AI models are built. It also disregards the fact that “destruction or other reasonable disposition” and impounding of infringing copies are long-established statutory remedies for copyright infringement that courts can (and do) order, notwithstanding the economic impact on the infringer.⁵⁷ We have already seen the disastrous impact on the creative community stemming from technology companies that built their businesses using valuable copyrighted works without permission, and we should not have to relive those days here.

Turning to the substance of the question, machine unlearning is “an emergent subfield of machine learning that aims to remove the influence of a specific subset of training examples – the ‘forget set’ – from a trained model.”⁵⁸ Although it remains to be seen whether an AI model can reliably “unlearn” what it gained from training on a particular piece of training material, there is a marketplace need for this capability (such as to remove personally identifiable information and other privacy breaches, eliminate biased data from datasets, or remove copyrighted works used without permission) that will ensure that such tools are developed and commercialized.⁵⁹ See section 4 of Annex A for various articles/approaches on machine unlearning.

Relatedly, the Federal Trade Commission (FTC) has begun using “algorithmic disgorgement” as an enforcement tool and has included it in a number of recent settlements.⁶⁰ Also referred to as “model deletion,” the enforcement strategy requires companies to delete products built on data they should not have used in the first place. For instance, if the FTC finds that a company trained a large language model on improperly obtained data, then it will have to delete all the information along with the products developed from the ill-gotten data.”⁶¹ Since 2019, the FTC has used this remedy more than once under its “broad authority to fashion appropriate remedies for violations of the [Federal Trade Commission] Act.”⁶²

7.4. Absent access to the underlying dataset, is it possible to identify whether an AI model was trained on a particular piece of training material?

⁵⁷ See 17 U.S.C. § 503.

⁵⁸ Fabian Pedregosa & Eleni Triantafyllou, *Announcing the first Machine Unlearning Challenge*, Google Research Blog (June 29, 2023); see also Than Tam Nguyen et al., *A Survey of Machine Unlearning*, arXiv, Oct. 21, 2022, available at <https://arxiv.org/pdf/2209.02299v5.pdf>.

⁵⁹ See Tom Simonite, *Now That Machines Can Learn, Can They Unlearn?*, WIRED (Aug. 19, 2021, 7:00 AM), <https://www.wired.com/story/machines-can-learn-can-they-unlearn/> [https://perma.cc/VB9G-BGN8],

⁶⁰ See, e.g., Decision and Order, In re Everalbum, Inc., Comm’n File No. 1923172 (FTC 2021), https://www.ftc.gov/system/files/documents/cases/1923172_-_everalbum_decision_final.pdf [https://perma.cc/ZKM4-H7Y5]; Stipulated Order, United States v. Kurbo Inc., No. 22-CV-00946 (N.D. Cal. 2022), https://www.ftc.gov/system/files/ftc_gov/pdf/wwkurbostipulatedorder.pdf [https://perma.cc/KKE9-6CAE].

⁶¹ Tony Riley, *The FTC’s Biggest AI Enforcement Tool? Forcing Companies to Delete Their Algorithms*, Cyberscoop (July 5, 2023), <https://cyberscoop.com/ftc-algorithm-disgorgement-ai-regulation>.

⁶² F.T.C. v. Pantron I Corp., 33 F.3d 1088, 1102 (9th Cir. 1994), referencing 15 U.S.C. §53(b).

Researchers are working on tools that allow them to analyze outputs generated by AI tools to determine specific inputs that were used to train the model.⁶³ See section 1 of Annex A for various articles/approaches on how to identify training material when the AI developer refuses to disclose it. At present, the work on these tools is most advanced with respect to text, including song lyrics, and visual images.⁶⁴ Efforts to do the same thing with speech are also underway. Although we are not aware of any commercial tools that are currently available to do this with respect to recorded music, it is likely just a matter of time until such tools are developed. As with machine unlearning, there is a marketplace need for these sorts of tools that will drive their development and commercialization.

8. Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.

Although we recognize that fair use involves a fact-specific, case-by-case analysis, the unauthorized reproduction of copyrighted works by AI developers to develop models that produce AI-generated works that actually or potentially compete with the inputted works comes as close as a use can come to being presumptively *not* fair use. As explained more fully below, the arguments of fair use that we frequently hear as justification for these practices, including references to the *Google Books* case, are misplaced. Unlike the *Google Books* case, where Google scanned millions of books without permission – not to generate new books, but to create a search function that pointed users to legitimate online locations of the copied books⁶⁵ – AI models are designed to use existing copyrighted works to create new expressive works that compete with or serve as substitutes for the originals. When such copying is done without authorization, that is not fair use, as the Supreme Court’s recent decision in the *Andy Warhol Foundation v. Goldsmith* case makes even clearer than it already was.⁶⁶

8.1. In light of the Supreme Court’s recent decisions in *Google v. Oracle America* and *Andy Warhol Foundation v. Goldsmith*, how should the “purpose and character” of the use of copyrighted works to train an AI model be evaluated? What is the relevant use to be analyzed? Do different stages of training, such as pre-training and fine-tuning, raise different considerations under the first fair use factor?

⁶³ See Ben Dickson, *Machine learning: What are membership inference attacks?*, TechTalks (Apr. 23, 2021) Available at <https://bdtechtalks.com/2021/04/23/machine-learning-membership-inference-attacks/>.

⁶⁴ See Section 1 of Annex A for articles that discuss this issue.

⁶⁵ *Authors Guild v. Google*, 804 F.3d 202 (2d Cir. 2015).

⁶⁶ The recent decision in *Thomson Reuters v. Ross Intel., Inc.* (No. 1:20CV613-SB, 2023 U.S. Dist. LEXIS 170155 (D.Del. Sept. 25, 2023)) is not to the contrary. That case, which held that the question of fair use in the particular circumstances of that dispute was a jury question, turned on the fact that the copyrighted works at issue were “far from the core of copyright” and therefore involved different considerations from those to apply to the use of creative works, such as sound recordings, for training of AI systems. See also, e.g., Katherine Lee et al., *Talkin’ ‘Bout AI Generation: Copyright and the Generative-AI Supply Chain*, arXiv, Sept. 15, 2023, at 99-100, available at <https://arxiv.org/ftp/arxiv/papers/2309/2309.08133.pdf>.

Of the two cases cited above, the *Warhol* case is far and away the most important and the most relevant. Not only does it post-date the *Google v. Oracle* decision, making it the Court’s latest pronouncement on matters of fair use, but the facts and the reasoning of the *Warhol* case are far more germane to the AI context. *Google v. Oracle* was a case about functional software code that the Court described as “further than are most computer programs . . . from the core of copyright.”⁶⁷ Although AI systems *are* software, the focus of this study is on the ingestion of creative copyrighted works, the creation of AI models, and the generation of new, competing works *by* the AI developer through their AI software.

After *Warhol*, the key inquiry under the first fair use factor is whether the purpose of the AI tool is to generate content that actually or potentially competes with the ingested work(s).⁶⁸ That is precisely what unlicensed generative AI tools do – they reproduce copyrighted works, without the rightsholder’s permission, in order to generate new material that competes in the marketplace with the very works that were used to build the system.⁶⁹ As noted above, this is not like the *Google Books* case that AI developers like to cite. There, the search engine served a wholly different purpose from the copyrighted works on which it was built. With generative AI tools, by contrast, the purpose for which unlicensed copyrighted works are ingested is to generate expressive works that are plainly substitutional. Under such circumstances, the use of unlicensed copyrighted works to develop an AI model is not transformative. We address the “commercial nature” or “nonprofit educational purposes” of the use in response to Question 8.3 below but note here the *Warhol* Court made clear that that aspect of the first fair use factor analysis “is not dispositive,”⁷⁰ nor should it be.

To answer the second question, the relevant uses to be analyzed are any points in the AI generation process in which copyright infringement occurs, whether that is the unauthorized copying that occurs in the creation and ingestion of the training dataset, the creation of an unauthorized derivative work in the form of an AI model that contains representations of some or all of the copyrighted material ingested for development purposes, which we discuss more fully in response to Question 14, or the outputs of the model that contain copyrighted elements of the ingestion material, see discussion of Section 114(b) in response to Question 2.

To answer the third question, different stages of development, such as pre-training and fine-tuning do not raise different legal considerations. In either case, the copying would be considered unauthorized and infringing. It will likely be easier to establish the connection in the case of a fine-tuned model because, in most instances, the universe of works used to fine-tune the model will be far smaller and narrower than the dataset used for pre-training.

⁶⁷ *Google LLC v. Oracle America, Inc.*, 141 S. Ct. 1183, 1202 (2021).

⁶⁸ *Warhol*, 598 U. S. ____ (2023) slip op. at 19 (“An independent justification like . . . [the need to copy in order to serve a new purpose] is particularly relevant to assessing fair use where an original work and copying use share the same or highly similar purposes, or where wide dissemination of a secondary work would otherwise run the risk of substitution for the original or licensed derivatives of it.”).

⁶⁹ *Warhol*, slip op. at 15 (“the first factor relates to the problem of substitution—copyright’s *bête noire*. The use of an original work to achieve a purpose that is the same as, or highly similar to, that of the original work is more likely to substitute for, or “supplan[t],” the work.” (citation omitted)).

⁷⁰ *Id.* at n.6.

8.2. How should the analysis apply to entities that collect and distribute copyrighted material for training but may not themselves engage in the training?

As discussed in response to Question 25, copyright liability is joint and several – if there is a shared purpose or common enterprise, the different entities are still joint tortfeasors, and all are liable for copyright infringement. If the purpose of creating a training set was to use it for subsequent training, then the entity collecting and distributing copyrighted material for ingestion and the one doing the ingestion are jointly and severally liable.

8.3. The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes. How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature? Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

See answer to Question 8.1. At the outset, we wish to emphasize that the courts cannot allow actual or purported noncommercial/research use to function as a free pass. Even if a court were to view an entity’s noncommercial or research purposes⁷¹ as relevant to a fair use analysis, that is only one of four fair use factors, each of which must be separately considered and “weighed along with other factors.”⁷² Furthermore, as the *Warhol* Court made clear, “whether [a] use is of a commercial nature or is for nonprofit educational purposes’ is one element of the first factor, §107(1); it does not dispose of that factor, much less the fair use inquiry.”⁷³ Indeed, even noncommercial or nonprofit uses can weigh against fair use under the first factor if the uses result in market harm or the accrual of benefits to the alleged infringer.⁷⁴

Turning to the second part of the question, the *Warhol* case specifically teaches that “each challenged use must be assessed on its own terms.”⁷⁵ If a use starts out as noncommercial or for

⁷¹ It bears remembering, of course, that Section 107(1) of the Copyright Act is concerned with the “purpose and character of the *use*, including whether such *use* is of a commercial nature or is for nonprofit educational *purposes*” (emphases added); not with the commercial or noncommercial status of the *entity*. It is essential that the courts not conflate the nature of the use with the status of the entity making the use, as many non-profit entities, such as universities and research institutes, engage in for-profit, commercial activities. See, e.g., *Hachette v. Internet Archive*, ___ F. Supp. 3d ___, No. 20-cv-4160 (JGK), 2023 U.S. Dist. LEXIS 50749 (S.D.N.Y. 2023) (where an entity accrued benefits from an allegedly noncommercial use that it funded that weighed against a finding of fair use under the first factor).

⁷² *Warhol*, slip op. at 18, n.6 (citation omitted).

⁷³ *Id.* at 25, n.13.

⁷⁴ See, e.g., *Hachette Book Grp., Inc. v. Internet Archive*, ___ F. Supp. 3d ___, 2023 U.S. Dist. LEXIS 50749 (S.D.N.Y. 2023) (scanning of print books and digitally lending them without authorization is not fair use); *Worldwide Church of God v. Phila. Church of God, Inc.*, 227 F.3d 1110, 1118 (9th Cir. 2000) (copying by a nonprofit religious organization of an out-of-print book is not fair use); *Soc’y of Holy Transfiguration Monastery, Inc. v. Gregory*, 689 F.3d 29, 61 (1st Cir. 2012) (posting of plaintiff’s English translations of ancient religious texts on defendant Archbishop’s website did not constitute fair use); *Blackwell Publ’g, Inc. v. Excel Research Grp., LLC*, 661 F. Supp. 2d 786 (E.D. Mich. 2009) (copying by third party of copyrighted content in course packs without authorization to sell them to students was not fair use); *Am. Geophysical Union v. Texaco, Inc.*, 60 F.3d 913 (2d Cir. 1995) (copying of articles by researchers beyond what was permitted in the license for the articles was not fair use).

⁷⁵ *Warhol*, slip op. at 6 (Gorsuch, J., concurring) (“[W]hile the Foundation may often have a fair-use defense for Mr. Warhol’s work, that does not mean it always will.”).

nonprofit educational purposes but is later adapted for use of a commercial nature, the latter is the decisive factor for purposes of applying the fair use analysis.

Lastly, where funding is provided by for-profit developers for so-called noncommercial or research uses, that should be viewed as a red flag and should weigh heavily against any finding that the first fair use factor favors the uses that purport to be noncommercial or for research purposes. Given the degree of disaggregation within the generative AI supply chain, it is important that courts take such red flags seriously. Otherwise, were they to presumptively treat uses made by supposedly noncommercial entities differently than uses made by commercial entities for purposes of the first fair use factor, they would promote gamesmanship and data laundering. *See* response to Question 9.1.

8.4. What quantity of training materials do developers of generative AI models use for training? Does the volume of material used to train an AI model affect the fair use analysis? If so, how?

As we understand it, the quantity of copyrighted works ingested by generative AI models varies widely, depending on the nature of the model and the purpose for which it is being trained. Although AI developers often claim that they need unfettered access to the entire existing corpus of creative works, this is not true. Very powerful AI tools can be built with less data than some AI advocates contend. Adobe, for example, has done so with its Firefly AI tool, which Adobe claims it built by using only proprietary, licensed, or public domain materials as training data.⁷⁶ Indeed, models that limit themselves to a smaller universe of high-quality inputs may create higher quality outputs.⁷⁷ And models that use a smaller universe of inputs typically require fewer licenses.

Whether a particular model ingests a massive amount of copyrighted content or a more modest amount, the legal analysis is the same. Unless a specific use qualifies for fair use after all four factors are properly analyzed and balanced, it is infringing to reproduce or distribute copyrighted works (whether the number of works is small or a large) without permission of the copyright owner. To argue, as many in the AI community do, that *massive* infringement is somehow permissible fair use while more *limited* infringement is not would turn established principles of copyright law on their head. If anything, the volume of material used should make a finding of fair use less likely, as the larger the volume of works that is used without permission, the wider the swath of the market that is impacted by the unauthorized use and the greater the harm to the licensing market for AI model development. *See* response to Question 8.5.

⁷⁶ Adobe, <https://www.adobe.com/sensei/generative-ai/firefly.html#:~:text=Trained%20on%20Adobe%20Stock%20images,dataset%20to%20retrain%20Firefly%20models> (last visited on Oct. 23, 2023) (claiming that Adobe Firefly is “[t]rained on Adobe Stock images, openly licensed content, and public domain content...”).

⁷⁷ Matt Mullen, *AI music wars: Meta takes on Google and releases its own AI music generator – but whose is better?*, MusicRadar (June 16, 2023), <https://www.musicradar.com/news/meta-google-ai-music-wars-musicgen> (last visited Oct. 23, 2023) (concluding that Meta’s product, which is trained on significantly less music than Google’s product, creates better music). *See also* Jade Copet et al., *Simple and Controllable Music Generation*, arXiv, June 8, 2023, available at <https://arxiv.org/pdf/2306.05284.pdf>.

To the extent that they are relevant to this question, the arguments that AI developers like to make about the difficulty of obtaining licenses for all of the copyrighted inputs they want or need are misplaced when it comes to recorded music. As described in response to Question 2, our members have more than amply demonstrated that free-market licensing of sound recordings is feasible and adaptable to the new world of generative AI.

8.5. Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured? Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?

The legal standard to be applied under the fourth fair use factor is clear. The inquiry should be whether “if the challenged use ‘should become widespread, it would adversely affect the potential market for the copyrighted work.’”⁷⁸ In an infringement action involving copyrighted works ingested into an AI model, the court should consider the effects of the type of conduct involved on all potential markets for the work(s) in suit, including (but not limited to) competition with those works by outputs of that model as well as the harm to the licensing market for the works, which includes licensing the works for distribution⁷⁹ as well as licensing the works for AI model development. For example, a market for licensing the use of copyrighted works in AI models is emerging,⁸⁰ and harm in that market should be taken into account.⁸¹ Moreover, the court should not limit its assessment of market harm to the particular model involved but should also consider the potential effects of widespread development and use of other, similar models incorporating the works in suit. Today, our members are in the business of licensing their content; any unauthorized use that interferes with actual or potential licensing markets necessarily adversely affects their core business. Moreover, unlicensed use of our members’ content for AI training will negatively impact the existing streaming marketplace (as unauthorized AI generated content competes directly with our existing content).

9. Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?

The Copyright Act establishes an opt-in, permissions-based regime. Section 106 grants copyright owners certain exclusive rights; those rights include the right to authorize (or not authorize) others to exercise those same rights. There is no basis in law or policy for imposing an opt-out regime.

9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?

⁷⁸ Harper & Row, Publs. v. Nation Enters., 471 U.S. 539, 568 (1985) (quoting Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 451 (1984)); *see also* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 590 (1994).

⁷⁹ Complaint & Demand for Jury Trial at ¶ 46, Concord Music Group, Inc. v. Anthropic PBC, No. 3:23-cv-01092 (M.D. Tenn. Oct. 18, 2023) (For example, music publishers have a market for licensing their song lyrics to lyrics aggregators and other websites that give consumers access to authorized copies of the lyrics to their favorite songs).

⁸⁰ *See supra* note 45.

⁸¹ *See, e.g.,* Am. Geophysical Union v. Texaco Inc., 60 F. 3d 913, 930 (2d Cir. 1994) (“[A]n impact on potential licensing revenues for traditional, reasonable, or likely to be developed markets should be legally cognizable...”).

We take issue with the framing of the question as it suggests that consent of the copyright owner is not already required for all uses, including noncommercial uses.⁸² As noted above, copyright law is an opt-in, permissions-based system. Accordingly, consent of the copyright owner is required for any and all reproductions and other exclusive uses of copyrighted works, whether the use is for noncommercial or research purposes or for avowedly commercial, for-profit purposes. The only exception is where the AI developer can meet its burden to establish that the reproductions and other uses are fair use under the particular circumstances, weighing all of the fair use factors in accordance with the Supreme Court’s guidance, as most recently articulated in the *Warhol* case. Any other approach would encourage gamesmanship and data laundering as for-profit, commercial entities seek to cloak the collection and reproduction of massive amounts of copyrighted works under the rubric of noncommercial/research activities or vest those particular responsibilities in noncommercial/research entities. *See* response to Question 8.3.

9.2. If an “opt out” approach were adopted, how would that process work for a copyright owner who objected to the use of their works for training? Are there technical tools that might facilitate this process, such as a technical flag or metadata indicating that an automated service should not collect and store a work for AI training uses?

First and foremost, an opt-out approach would require a fundamental change in U.S. copyright law, which is and always has been a permissions-based, opt-in regime. Moreover, the opt-in licensing approach works well in the recorded music business and there is no basis for altering that approach.⁸³ Our members have been licensing their full catalogs to Digital Music Services for decades, and those licenses have created a thriving, digital music business. The infrastructure exists for AI developers to engage with rightsholders to obtain the licenses they need on an “opt-in” basis instead of undertaking infringing activities. There is no justification for AI developers to proceed without obtaining the necessary licenses.

In addition, an opt-out system would be fundamentally unfair to creators and copyright owners. As has been well-documented elsewhere,⁸⁴ these same creators and copyright owners already spend far too much of their time and resources playing whack-a-mole with online service providers pursuant to Section 512’s unwieldy notice-and-takedown system. Expecting those same creators and copyright owners to now patrol AI datasets and models, most of which lack

⁸² *See supra* note 74.

⁸³ This is true, notwithstanding the fact that some very specific uses of sound recordings are subject to statutory licenses that impose very strict terms and conditions on the parties utilizing those licenses. Indeed, many of the services that used to rely on the Section 114 statutory licenses have migrated to voluntary licenses instead, because they offer expanded functionality and added flexibility. The economic success that our members have derived from negotiating voluntary marketplace deals with a plethora of Digital Music Services that offer different functionality and different business models makes clear that there is no need to expand the use of statutory licenses in our industry.

⁸⁴ *See* U.S. Copyright Office, Section 512 of Title 17, at 33 (2020), <https://www.copyright.gov/policy/section512/section512-full-report.pdf> (“[T]he volume of notices demonstrates that the notice-and-takedown system does not effectively remove infringing content from the internet; it is, at best, a game of whack-a-mole.”).

the transparency needed to enable proper enforcement, to first locate their works and then file an endless succession of opt-out requests would be both infeasible and unfair.⁸⁵

9.3. What legal, technical, or practical obstacles are there to establishing or using such a process? Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?

For the first question, *see* response to Question 9.2. For the second question, *see* response to Question 8.4 above.

From a legal perspective, as mentioned above, U.S. copyright law is an opt-in system. From a technical perspective, flags that copyright owners include will not be effective unless developers also create and deploy tools that can accurately find, recognize, and comply with some sort of “do not train” flag. From a practical perspective, it is unlikely that an opt-out process would work unless AI developers faced significant penalties for non-compliance. *See* response to Question 9.4. In addition, there is a proliferation of different technical standards for opt-out flags, including propriety flags promoted by individual AI developers. It would be burdensome, unreasonable, and ineffective for copyright owners to have to tag their content with a multiplicity of different tags simply to prevent their copyrighted works from being copied and ingested without their permission. This would be so even if copyright were an opt-out system; it is doubly so given that copyright is an opt-in system.

The volume of works used in developing an AI model or tool makes it more imperative, not less, that AI developers obtain consent in advance from copyright owners. When building a business that relies on others’ copyrighted works, obtaining consent is a necessary cost of doing business. That said, the fact that generative AI tools can be built with less data than many AI advocates like to claim, *see* response to Question 8.4, offers AI developers a mechanism for exercising some control over the number of licenses they need to negotiate. Most importantly, as noted repeatedly in these comments, the recorded music industry has all the necessary systems and infrastructure already in place to make obtaining advance consent demonstrably feasible.

9.4. If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?

As mentioned above, we oppose an opt-out approach, which is both contrary to U.S. law and unjustifiable in our industry. That said, if opt-out tools are developed and put into mainstream use, and if creators and copyright owners choose to use them to *augment* their existing protections, there must be strong remedies against AI developers and systems that ignore or fail to recognize those tools. At a minimum, if an AI developer ignores an opt-out technical measure associated with a copyrighted work and ingests that work anyway, that should be considered willful infringement and subject to a heightened award of statutory damages. In addition, depending upon how such an opt-out measure is implemented, it may also qualify as copyright

⁸⁵ *See, e.g.,* Kali Hays, “OpenAI offers a way for creators to opt out of AI training data. It’s so onerous that one artist called it ‘enraging,’” Business Insider, Sep. 29, 2023, available at <https://www.businessinsider.com/openai-dalle-opt-out-process-artists-enraging-2023-9>.

management information as that term is defined in in Section 1202(c)⁸⁶ As such, removal of the opt-out measure would constitute a violation of Section 1202(b), in addition to any other remedies that might be available.

To answer the second question, provided the failure to comply with opt-out technical measures is treated as willful infringement under Section 504 of the Copyright Act, existing remedies are adequate and no separate cause of action is needed.

9.5. In cases where the human creator does not own the copyright—for example, because they have assigned it or because the work was made for hire—should they have a right to object to an AI model being trained on their work? If so, how would such a system work?

Generally, a human creator who does not own the copyright in a work does not have a right to bring an infringement claim. However, in some cases, a human creator who assigned rights to a copyrighted work in exchange for an ongoing royalty may be considered a beneficial owner of the work and have standing to sue under Section 501(b) if the use at issue was not authorized by the actual copyright owner.⁸⁷ Also, as discussed primarily in response to Questions 2, 30, and 31, human creators should have robust rights to challenge unauthorized uses of their name, image, voice, and likeness.

10. If copyright owners’ consent is required to train generative AI models, how can or should licenses be obtained?

Once again, we take issue with the way the question is framed. Copyright owners’ consent *is* required to train generative AI models. With respect to licensing, we are strong proponents of voluntary marketplace agreements, and our industry has demonstrated that free-market licensing of sound recordings works well. As Jeff Harleston, the General Counsel and Executive Vice President of Business and Legal Affairs of Universal Music Group, said in written testimony submitted to the Senate Judiciary Committee: “We have a robust free market for sampling, sync licensing, deals with new entrants to the digital marketplace, social media companies and all manner of new technologies.”⁸⁸ As described more fully in response to Question 2, that market is the foundation of today’s thriving digital music ecosystem, in which all manner of Digital Music Services have negotiated licenses with our member companies and from independent labels directly or via distributors or organizations such as Merlin. Regarding how licenses can or should be obtained, *see* our response to Question 6.2.

⁸⁶ See 17 U.S.C. § 1202(c)(6) (copyright management information includes “[t]erms and conditions for use of the work.”).

⁸⁷ See, e.g., *Cortner v. Israel*, 732 F.2d 267, 271 (2d Cir.1984).

⁸⁸ See *Artificial Intelligence and Intellectual Property – Part II: Copyright: Before the S. Comm. On the Judiciary, Subcomm. Intell. Prop.*, 3 (2023) (written testimony of Jeff Harleston, General Counsel and Executive Vice President of Business and Legal Affairs, Universal Music Group), https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_harleston1.pdf.

10.1. Is direct voluntary licensing feasible in some or all creative sectors?

We cannot speak for other creative sectors, but direct voluntary licensing is demonstrably feasible in the recorded music sector. *See* responses to Questions 2, 6.2, and 10.

10.2. Is a voluntary collective licensing scheme a feasible or desirable approach? Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

Voluntary collective licensing that happens in the free market, without any government mandate or intervention, can be both desirable and feasible, as exemplified by the success of the digital rights agency Merlin. Voluntary marketplace negotiations are *always* preferable to government-imposed solutions. Competitive markets result in better products and services, as well as increased choices for consumers, generally on a faster timeline. To foist a new government-mandated collective licensing scheme on an industry like ours, that has already proven it does not need one, would be unfair, inappropriate, and damaging to the marketplace.

In answer to the last question, we are not opposed to antitrust exceptions if used to facilitate voluntary marketplace negotiations, such as those contemplated in the recently introduced Protect Working Musicians Act of 2023, H.R. 5576.⁸⁹ This would leave licensing in the free market while allowing smaller entities, which could not otherwise negotiate successfully against giant tech companies to enjoy a level playing field when negotiating with dominant streaming platforms and AI developers.

10.3. Should Congress consider establishing a compulsory licensing regime? If so, what should such a regime look like? What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out? How should royalty rates and terms be set, allocated, reported and distributed?

Congress absolutely should *not* consider establishing a compulsory licensing regime for recorded music. As an industry that has decades of first-hand experience with compulsory licensing, we are more familiar than most with the limits of a compulsory licensing regime, including below-market royalty rates, additional administrative costs, and most importantly, restrictions on innovation. Currently, the real economic engine for the digital music ecosystem and source of rapid growth in recent years is the portion of the market that operates free of statutory licenses.⁹⁰ Most important, the market for voluntary, full catalog sound recording licenses is working. *See* responses to Questions 2, 6.2, and 10.

⁸⁹ Protect Working Musicians Act of 2023, H.R. 5576, 118th Cong. (2023), <https://www.govinfo.gov/app/details/BILLS-118hr5576ih>.

⁹⁰ 94% of revenues for the first half of 2023 came from free-market, non-statutory royalties.

10.4. Is an extended collective licensing scheme a feasible or desirable approach?

No. There is no need or basis for government intervention in the licensing market for recorded music. The market is demonstrably working. *See* response to Question 10.3 above.

10.5. Should licensing regimes vary based on the type of work at issue?

We can only speak for our industry where voluntary marketplace licensing is already happening and should be permitted to flourish.

Were Congress to intervene in other industries, we caution it to do so in a way that it is strictly limited to only those industries that have either themselves requested Congressional intervention or where there has been a demonstrable need based on hard, economic data and other evidence of market failure. Where a need has been demonstrated, we urge Congress to begin by granting antitrust exemptions, if necessary, to facilitate voluntary licensing and give those time to work before considering any form of government-mandated licensing.

11. What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training? Who, if anyone, should be responsible for securing them (for example when the curator of a training dataset, the developer who trains an AI model, and the company employing that model in an AI system are different entities and may have different commercial or noncommercial roles)?

In the recorded music sector, there are no legal, technical, or practical issues that act as impediments with respect to obtaining appropriate licenses to reproduce, ingest, and otherwise exploit our member companies' sound recordings. *See* responses to Questions 2, 6.2, and 10.

When considering who, if anyone, should be responsible for securing licenses that cover reproduction and ingestion of our member companies' sound recordings, any entity (regardless of their identified role) that intends to exploit our members' copyrighted sound recordings needs a license from the copyright owner (or an express sublicense from a licensed intermediary). To put it another way, any entity whose actions would constitute infringement if done without a license must obtain a license to engage in the otherwise infringing activities. Disaggregating the supply chain does not change the analysis or eliminate liability. Anyone engaged in the AI "supply chain" that is exercising any of the copyright owner's Section 106 rights must obtain a license before doing so.

As discussed more fully above, in this context, we see no legally significant difference between commercial and noncommercial entities and reiterate the concerns about gamesmanship and data laundering raised in response to Questions 8.1, 8.3, and 9.1.

12. Is it possible or feasible to identify the degree to which a particular work contributes to a particular output from a generative AI system? Please explain.

The academic field in this area is still nascent but quite active. *Please see* section 2 of Annex A for a subset of papers setting out research which seems potentially applicable to the analysis and

estimation of the contribution of training data influence on outputs of a generative AI system. These papers suggest there may be some avenues to determine or estimate the influence of training data on the outputs in some generative AI systems.⁹¹ However, to our knowledge, such techniques are not commercially available, and it is currently not possible to effectively inspect AI models and determine the influence of particular training data at scale. This is because such systems do not appear to be designed with the aim of allowing the influence of specific inputs to be tracked. If that were a design imperative, we believe it would be possible for an AI model to be built in such a way that attribution of the influence of inputs is part of the core model. Increasingly there is demand in other fields that AI cannot be a black box and must be 'Explainable AI,' capable of demonstrating the basis on which its decisions or outputs were generated.⁹² This principle should also apply to the use of copyrighted materials.

Moreover, these approaches assume that the inspector has access to the training data and the model, which is often not the case for those other than the AI developer. While AI developers should have the primary responsibility to understand how their model operates, others, such as rightsholders and regulators, may also need to inspect and analyze AI models. Given this, it is essential that rightsholders and regulators have greater transparency into training datasets and AI model attributes to better understand the models and their attendant risks.

13. What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?

Once again, we take issue with the implied premise of the question. As noted above, U.S. copyright law is an opt-in system, which means that a licensing requirement already exists for the development and adoption of generative AI systems. Unfortunately, the question reflects a familiar (and unfounded) complaint that we hear from entities that seek to build their businesses on the backs of copyright owners; i.e., that licensing is too difficult or costly. The flourishing market for Digital Music Services – which is built entirely on content licensed in the free market – shows that licensing can give rise to a thriving market that lifts all boats. Moreover, voluntary licenses allow AI developers to obtain global rights from respective recorded music rightsholders, typically in a single negotiation. This helps AI developers avoid the need to navigate a multiplicity of different and changing legal regimes across the globe.

When considering economic impacts, a better question to ask is what would be the economic impacts of a market built on piracy and infringement?⁹³ Without licenses in place, AI developers would face serial lawsuits from content owners large and small, requiring large outlays of money

⁹¹ One of the papers also notes the high influence of famous passages in large language models, which has led to retrieval of those passages in the output of the model. See Roger Grosse et al., *Studying Large Language Model Generalization with Influence Functions*, arXiv, Aug. 7, 2023.

⁹² See, e.g., Ricardo Guidotti et al., *Local Rule-Based Explanations of Black Box Decision Systems*, arXiv, May 28, 2018, available at <https://arxiv.org/pdf/1805.10820.pdf>; Patrick Rehill & Nicholas Biddle, *Transparency Challenges in Policy Evaluation with Casual Machine Learning – Improving Usability and Accountability*, arXiv, Oct. 20, 2023, available at <https://arxiv.org/pdf/2310.13240v1.pdf>; Syzmon Bobek et al., “*Local Universal Rule Based Explanations*,” arXiv, Oct. 23, 2023, available at <https://arxiv.org/pdf/2310.14894.pdf>.

⁹³ This is consistent with the fourth of the four fair use factors, which asks about “the effect of the use upon the potential market for or value of the copyrighted work.” 17 U.S.C. § 107(4). Nowhere in the fair use analysis are courts required to consider the potential impact on the user or the user’s market. That is as it should be.

for outside lawyers and large outlays of time for in-house staff. This is not conjecture – consider the number of lawsuits that already have been filed since ChatGPT burst into popular view in early 2023.⁹⁴ Moreover, in an unlicensed world, product development will be driven by defensive legal strategies, rather than by creative and entrepreneurial ideas.

14. Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models.

First, a significant factor in establishing liability is the ability to know what copyrighted works were included in the AI training set and ingested by the AI model. This is why transparency and recordkeeping, which are discussed more fully in the next section, are essential and must be mandated. There also must be strong remedies for failure to comply with any such requirements.

Second, the trained model, which stands between (a) the copyrighted works curated and ingested on the input side and (b) the creation of separate works on the output side, must not be overlooked when doing a copyright infringement analysis. Depending on the specific facts and circumstances, a trained model itself could be viewed as one or more of the following: an infringing derivative work, an infringing compilation, or an infringing reproduction. As a group of computer scientists at Cornell University explain:

One view is that a model is a compilation of its training data — the model is simply a different and complicated arrangement of training examples. Another view is that a model is a derivative work of its training data — “a work based upon one or more preexisting works . . . in which [those works are] recast, transformed, or adapted.” A derivative work (think of a translation of a novel, a recording of a song, or an action figure based on a character from a movie) combines the authorship in an existing (or ‘underlying’) work with new authorship. The substantive difference between the two is that in a compilation, the underlying works are present in substantially unmodified form, whereas in a derivative work the underlying work is ‘recast, transformed, or adapted.’⁹⁵

⁹⁴ See, e.g., *Chabon v. Meta Platforms, Inc.*, No. 4:23-cv-04663-DMR (N.D. Cal. Sept. 12, 2023); *Chabon v. OpenAI, Inc.*, No. 3:23-cv-04625 (N.D. Cal. Sept. 8, 2023); *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417 (N.D. Cal. Jul. 7, 2023); *P.M. v. OpenAI, Inc.*, No. 3:23-cv-03199 (N.D. Cal. June 28, 2023); *Tremblay v. OpenAI, Inc.*, No. 3:23-cv-03223 (N.D. Cal. June 28, 2023); *L. v. Alphabet Inc.*, No. 3:23-cv-03440 (N.D. Cal. July 11, 2023); *Silverman v. OpenAI, Inc.*, No. 3:23-cv-03416 (N.D. Cal. July 7, 2023); *Andersen v. Stability AI Ltd.*, No. 3:23-cv-00201-WHO (N.D. Cal. Jan. 13, 2023); *Getty Images (U.S.), Inc. v. Stability AI, Inc.*, No. 23-135 (D. Del. Feb. 3, 2023); *Authors Guild v. OpenAI Inc.*, No. 1:23-cv-8292 (S.D.N.Y. Sept. 19, 2023); *Concord Music Group Inc. v. Anthropic PBC*, (M.D. Tenn. Oct. 18, 2023); *Huckabee v. Meta Platforms, Inc.*, No. 1:23-cv-09152 (S.D.N.Y. Oct. 17, 2023). See also Christopher T. Zirpoli, Cong. Rsch. Serv., LSB10922, *Generative Artificial Intelligence and Copyright Law 4* (2023), <https://crsreports.congress.gov/product/pdf/LSB/LSB10922> (“Plaintiffs have filed multiple lawsuits claiming the training process for AI programs infringed their copyrights in written and visual works. These include lawsuits by the Authors Guild and authors Paul Tremblay, Michael Chabon, Sarah Silverman, and others against OpenAI; separate lawsuits by Michael Chabon, Sarah Silverman, and others against Meta Platforms; proposed class action lawsuits against Alphabet Inc. and Stability AI and Midjourney; and a lawsuit by Getty Images against Stability AI.”).

⁹⁵ See Katherine Lee et al., *Talkin’ Bout AI Generation: Copyright and The Generative-AI Supply Chain*, arXiv, at Sept. 14, 2023, at 54, available at <https://arxiv.org/ftp/arxiv/papers/2309/2309.08133.pdf>.

The important point is not to limit the infringement analysis to just the inputs and the outputs, but to also consider the model itself, which sits between those two endpoints.

Transparency & Recordkeeping

15. In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?

Yes (developers of AI models should be required to collect, retain, and disclose records regarding the materials used to train their models) and yes (creators of training datasets should have a similar obligation). Most AI systems are a virtual “black box,” as it is impossible to discern most aspects of their development and operation from the outside. Accordingly, to protect people’s rights and safety, including copyright rights, rights of publicity, rights of privacy, and other human rights, AI systems must be designed and deployed in a manner that is accountable, transparent, and trustworthy. This issue is much broader than copyright concerns, and includes addressing concerns about performance, quality of results, and elimination of bias, among others. Congress has already taken a step toward addressing these issues through the National Artificial Intelligence Initiative, a program to “lead the world in the development and use of trustworthy artificial intelligence systems,” including with respect to “ethical” and “legal” issues.⁹⁶

As relevant here, complete recordkeeping of copyrighted works, including the way in which they were used to develop and train any AI system as well as to generate particular outputs, is essential. To support this objective, U.S. policy should require AI developers and deployers to keep proper records, ensure that those records are periodically audited, and disclose appropriate information about their AI systems. Consistent with that principle, AI developers and deployers who use third-party training datasets or pre-trained models should obtain, retain, and make available necessary data from upstream sources. To avoid “AI laundering,” it is essential that any U.S. recordkeeping requirements apply to all systems that are made available in the U.S. or that generate output used in the U.S., regardless of the geographic location in which the development (including pre-training, training, fine-tuning, validation, testing, adapting a pre-trained AI system, or the generation of outputs) may have taken place. Even though individually tailored reporting provisions will be a key term of any voluntary licenses, the sort of recordkeeping proposed here and in response to Question 15.1 will expedite free-market licensing negotiations by providing copyright owners with an important baseline regarding which of their works have been ingested and how they have been used.

Making such recordkeeping legally required would help deter copyright infringement. However, because it can be nearly impossible (from the outside) to ascertain if and how copyright-protected content has been used to train AI models and whether infringements have taken place, it will be difficult if not impossible for copyright owners to monitor the unauthorized use of their works without access to granular records. Thus, the U.S. should consider providing a mechanism for those claiming to be harmed by the AI system to obtain or be able to inspect

⁹⁶ National Artificial Intelligence Initiative Act of 2020, Pub. L. No. 116-283, §§ 5101(a)(2), 5103(d)(2)(C), 134 Stat. 3388, 4523, 4527 (2021).

applicable records and audits of the AI system to determine whether a person’s rights have been infringed or otherwise violated. *See, for example*, our discussion of a potential new administrative subpoena process described in response to Question 15.2.

15.1. What level of specificity should be required?

As described above, copyright owners with a legitimate interest in knowing whether their works may have been infringed in the development of an AI system should have access to the most granular information necessary to make such a determination and to be able to enforce their rights when it appears their copyrights have been infringed.

At a minimum, proper recordkeeping should include complete documentation about (1) what materials were ingested to develop an AI system (or to fine-tune or adapt a pre-trained AI system)⁹⁷ and in what manner, (2) the provenance of such materials, including whether any licenses or authorizations were sought or obtained to authorize such use and copies of those licenses or authorizations, (3) the articulated rationale for selecting and using the materials ingested for the AI system’s development, (4) the individual or organization responsible for the AI system (including who is responsible for ingesting the materials, who is responsible for any foundational AI model, who is responsible for any fine-tuning of the AI model, who is deploying the AI system, etc.) and (5) the extent to which each copyrighted work in the training materials was relied upon in order to generate specific outputs from the AI model.⁹⁸

For transparency, AI developers and deployers should be required to disclose to the public the purpose of the AI system and its overall functionality, who is the individual or entity responsible for the AI system and their location and contact information, the provenance of the materials ingested during the AI system’s development, and basic information to provide algorithmic transparency.⁹⁹

⁹⁷ This information should be granular enough that a rights owner can review the records and determine if their works have been ingested. For recorded music, that means that AI developers would have to track fields including, but not necessarily limited to, track title, artist name, songwriter name, ISRC, ISWC P-line, and C-line.

⁹⁸ As noted by the National Institute of Standards and Technology (NIST), “[m]aintaining the provenance of training data and supporting attribution of the AI system’s decisions to subsets of training data can assist with both transparency and accountability.” Nat’l Inst. of Standards and Tech., Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, 16 (2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

⁹⁹ This type of transparency is consistent with the *White House Blueprint for an AI Bill of Rights*, The White House, <https://www.whitehouse.gov/ostp/ai-bill-of-rights/> (last visited Oct. 24, 2023), and congressional calls for disclosure requirements for AI systems, such as those expressed at the Senate Judiciary Subcommittee on Privacy, Technology and the Law Hearing titled Oversight of A.I.: Rules for Artificial Intelligence, May 16, 2023 (a video of which is available at <https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence>) and the House Judiciary IP Subcommittee hearing titled Artificial Intelligence and Intellectual Property: Part I: Interoperability of AI and Copyright Law, May 17, 2023 (a video of which is available at https://www.youtube.com/watch?v=MmINQ_Kqumw). Such requirements are also under discussion in Europe. *See* Amendments Adopted by the European Parliament on the Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) Amdt. 399 (June 14, 2023), available at https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html. The pending EU AI Act, which has been approved by the European Parliament and now is the subject of a member state consultation process (which could result in changes to the current text), would require that providers of generative AI models “document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law.”

For more detail on the level of specificity that should be required, *please see* the June 12, 2023 Comments of the Creative Community on the National Telecommunications and Information Administration AI Accountability Policy Request for Comment (“Creative Community NTIA Comments”)¹⁰⁰ and the July 7, 2023 Comments of the Creative Community on the Office of Science and Technology Policy Request for Information on National Priorities for Artificial Intelligence (“Creative Community OSTP Comments”).¹⁰¹

15.2. To whom should disclosures be made?

As described above, information should be disclosed to regulators, interested parties, and the general public at a level of specificity necessary to address their legitimate concerns relating to the development, training, and operation of AI systems. Mandatory disclosure must also be done in a way that does not result in the unauthorized propagation of copyrighted works. This means that any mandated disclosure process must include appropriate provisions for keeping works confidential without stymying the purpose of the disclosure rules.

Copyright owners should also be able to demand that relevant portions of any records AI developers are required to keep (under proposed new recordkeeping laws) be disclosed to them. As noted in response to Question 5, one possibility for such a system is to create a new administrative subpoena process whereby a copyright owner or a person authorized to act on the owner’s behalf may, by asserting a good faith belief that one or more of the owner’s copyrighted works has been used by an AI developer without authorization from the owner, request the clerk of the United States district court to issue a subpoena to an AI developer for identification of any of the copyrighted works that have been reproduced or otherwise exploited by the AI developer. If the AI developer does not comply with any such subpoena, either because the developer has not kept the required records or because the developer chooses not to comply, then the copyright owner should be entitled to an evidentiary presumption, sufficient at least to support a copyright infringement lawsuit, that the copyright owner’s works were, in fact, reproduced. Without such a tool to compel disclosure and without serious sanctions for failure to comply with an administrative subpoena, any recordkeeping requirements that may be mandated will be meaningless. An outline of a specific legislative proposal can be found in Annex B hereto.

15.3. What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?

All parties in the generative AI process must have appropriate licenses (or permissible sublicenses) if they are reproducing or distributing copyrighted works (or exercising any of the other Section 106 exclusive rights). Parties involved in AI development cannot avoid liability by disaggregating the various steps in the “supply chain” (e.g., collecting copyrighted works, assembling a dataset, creating an AI model based on copyrighted works, creating an AI tool for third-party use, etc.). Where AI developers incorporate models from third parties, the AI developer must either obtain licenses directly from all affected copyright owners or obtain a

¹⁰⁰ Docket No. NTIA-2023-0005, available at <https://www.regulations.gov/comment/NTIA-2023-0005-1277>.

¹⁰¹ Docket No. OSTP-TECH-2023-0007. These comments are not yet available through Regulations.gov. A copy is attached as Annex C to these comments.

sublicense from the party that created the model, after using due diligence to confirm that the party that created the model has all rights necessary to grant such a sublicense.

Requiring all parties in the supply chain to have the necessary licenses/permissions is nothing novel. Newspapers (and other media businesses), for example, deal with this all the time. If a newspaper wants to publish a photograph that it acquires from a third party (e.g., a newswire, a stock photography company), it must ensure that its source obtained the photograph legally and that it is permitted to sublicense downstream users. Movie studios have teams of people that clear third-party works for inclusion in motion pictures and television shows. And our members have clearance departments that clear rights to third-party samples, artwork, and so forth. AI developers should not be treated any differently.

15.4. What would be the cost or other impact of such a recordkeeping system for developers of AI models or systems, creators, consumers, or other relevant parties?

Again, we take issue with the implied premise of the question. Keeping track of the copyrighted works that are used by AI systems or models is a necessary cost of using copyrighted works. Without such records, AI developers cannot adhere to many of the HAC's principles, which are set forth in the introduction to these Comments. Although the cost of recordkeeping should be fairly trivial (*see below*), even if that were not the case, the cost (or exaggerated claims about the cost) should not be a reason not to require AI developers to engage in reasonable recordkeeping. Recordkeeping is simply a cost of doing business.

The sort of recordkeeping proposed above should be easily automated, especially if the various entities involved in the AI supply chain ensure that the metadata that accompanies legitimately sourced sound recordings is maintained along with the sound recordings themselves. Indeed, there are already companies in the marketplace that provide services to help with and automate such recordkeeping.¹⁰² For example, an AI dataset or model that includes sound recording ISRCs should be able to easily identify the key data fields associated with each such sound recording and use that data to automatically populate a recordkeeping template. Similarly, AI developers could use audio fingerprinting technology to scan the sound recordings in the dataset, detect their titles and other relevant data fields, and automatically populate a recordkeeping template. Digital music services already have this technology, as they are required to provide monthly reports of use to record companies, the Mechanical Licensing Collective (the "MLC") (the collective management organization that administers the Section 115 compulsory license), and, for some digital music services, to SoundExchange (the collective management organization that administers the Section 114 statutory license). Likewise, YouTube, Meta, and other UGC-oriented platforms have developed tools that automatically scan all uploaded content, track it, and generate reports of use.

¹⁰² See, e.g., Marius Schlegel & Kai-Uwe Sattler, *Management of Machine Learning Lifecycle Artifacts: A Survey*, arXiv, Oct. 21, 2022, available at <https://arxiv.org/pdf/2210.11831.pdf> (reviews a representative selection of more than 60 systems and platforms).

16. What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?

The obligation to “notify” copyright owners that their works will be used to train an AI model already exists, except that notice must be provided in the form of a request for permission *before* a work can be used, not after works have been used, as the question assumes. As explained more fully in response to Questions 9 and 9.2, copyright is an opt-in, permission-based system. Expecting AI developers to obtain appropriate licenses before building their businesses on the back of other people’s creative works is nothing remotely new.

Each AI company should also be required to have its own reverse-search databases of copyrighted works (text, music, images, etc.) that copyright owners can use to determine if their works have been used to train that company’s AI model. They should also make available for public review a searchable database of URLs of webpages that have been scraped. In an ideal world, AI models should have the ability to disclose in response to an appropriate prompt (e.g., asking ChatGPT if it was trained on a particular work).

17. Outside of copyright law, are there existing U.S. laws that could require developers of AI models or systems to retain or disclose records about the materials they used for training?

Federal law includes many hundreds (possibly thousands) of recordkeeping and disclosure requirements, applicable to a myriad of activities and federal programs. Thus, applying similar requirements to developers of AI models or systems would not be out of the ordinary. In some specific cases, existing requirements might apply. For example, various federal government agencies fund AI research, and such research would be subject to recordkeeping and disclosure requirements in the agencies’ grant or contract regulations, as well as the terms of specific grants and contracts. For some projects, such requirements might involve records concerning training materials.

In addition, given statements made to date by members of Congress, the Administration, and those within the AI community,¹⁰³ there may be U.S. laws in the future that require developers of AI models or systems to retain or disclose records about the copyrighted materials they ingested.

Generative AI Outputs

If your comment applies only to a particular subset of generative AI technologies, please make that clear.

¹⁰³ See, e.g., House Science, Space and Technology Committee Hearing, *Trustworthy AI: Managing the Risks of Artificial Intelligence* (Sept. 29, 2022), <https://www.youtube.com/watch?v=BcdqyETo4Zg&t=3769s> (video of hearing); Senate Judiciary Subcommittee on Privacy, Technology and the Law Hearing, *Oversight of A.I.: Rules for Artificial Intelligence* (May 16, 2023), <https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence> (video of hearing); House Judiciary IP Subcommittee Hearing, *Artificial Intelligence and Intellectual Property: Part I: Interoperability of AI and Copyright Law* (May 17, 2023), https://www.youtube.com/watch?v=Mm1NQ_Kqumw (video of hearing).

Copyrightability

18. Under copyright law, are there circumstances when a human using a generative AI system should be considered the “author” of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?

As the Office and every court that has considered this question have consistently concluded, copyright only protects the unique value of human creativity.¹⁰⁴

With respect to text commands or prompts, we agree with the Office’s decision in the *Kashtanova* and *Allen* cases,¹⁰⁵ which both found that providing an iterative series of text commands or prompts was insufficient to claim authorship of the resulting AI output. While the text of those prompts may be independently copyrightable if sufficiently expressive, that does not confer upon the author of the prompt any copyright in the output generated by the AI system.

Lastly, we encourage the Office to keep in mind that, where an AI model is trained on unlicensed works, the model and its output are likely unauthorized derivative works of those unlicensed works. There can be no copyright in such unauthorized derivative works, even if they are further modified by human “collaborators.”

19. Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when content including AI-generated material is subject to copyright protection?

No. The case law and longstanding Office practice on the requirement of human authorship is clear. As the *Thaler*¹⁰⁶ court observed, every court to consider the issue has held that human authorship is a requirement for copyright protection.¹⁰⁷ The Office has confirmed the human authorship requirement in its registration decisions, including the *Kashtanova* case, the *Allen* case, and its March 16, 2023 registration guidance.¹⁰⁸ To the extent that the Office decides that additional standards are necessary to determine when content including AI-generated material is subject to copyright protection, those standards do not belong in the statute. Rather they belong

¹⁰⁴ See, e.g., *Thaler v. Perlmutter*, ___ F. Supp. 3d ___, 2023 WL 5333236, at*1 (D.D.C. Aug. 18, 2023) (“[H]uman authorship is an essential part of a valid copyright claim...”).

¹⁰⁵ Letter from Robert J. Kasunic, Associate Register of Copyrights and Director of Registration Policy and Practice, U.S. Copyright Office, to Van Lindberg, Re: Zarya of the Dawn 8-10 (Registration #Vau001480196), <https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>; Decision Re: Second Request for Reconsideration for Refusal to Register Théâtre D’opéra Spatial at 2 (SR # 1-11743923581; Correspondence ID: 1-5T5320R) (registration refused because “the deposit for the Work did not ‘fix only [Mr. Allen’s] alleged authorship’ but instead included ‘inextricably merged, inseparable contributions’ from both Mr. Allen and Midjourney”), <https://copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf>.

¹⁰⁶ *Thaler v. Perlmutter*, ___ F. Supp. 3d ___, 2023 WL 5333236, at *1 (D.D.C. Aug. 18, 2023).

¹⁰⁷ *Id.* See also, e.g., *Naruto v. Slater*, 888 F.3d 418 (9th Cir. 2018) (because Naruto was non-human, the monkey lacked standing under the Copyright Act); *Urantia Foundation v. Maaherra*, 114 F.3d 955, 958-59 (9th Cir. 1997) (“[I]t is not creations of divine beings that the copyright laws were intended to protect.”).

¹⁰⁸ Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190 (Mar. 16, 2023) (to be codified at 37 C.F.R. pt. 202).

in regulations, future registration guidance statements, Office circulars, the Compendium, and/or drop-down help menus in the registration applications included in the forthcoming Enterprise Copyright System.

20. Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

As explained in our response to Question 18, U.S. copyright law only protects works with human authorship and it should remain that way. We oppose the creation of new legal protection for AI-generated material. Indeed, as noted in response to Question 1, protection for works that lack human authorship would devalue human creativity and flood the marketplace with machine-made content that would make it harder for consumers to find human-made recordings from artists they love and support. Importantly, if the high volume of works that can be rapidly generated by AI tools were to become eligible for copyright protection (and registration), it would crowd out the field for true human creative expression and put human creators at enormous risk of inadvertently infringing one or more of the vast array of newly created, machine-made expressive works.

With respect to the second and third questions above, the staggering pace of investments currently being made in the development of generative AI systems suggests that there are already more than enough incentives in place to encourage the continued development of generative AI technologies and systems, notwithstanding clear statements from the Office and others that the material generated by such systems is not protectable.

20.1. If you believe protection is desirable, should it be a form of copyright or a separate sui generis right? If the latter, in what respects should protection for AI-generated material differ from copyright?

We do not believe that creation of new legal protection for AI-generated material is either warranted or desirable. We reiterate the risks identified in response to Question 20 that would be associated with any such protection.

21. Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material? Would such protection “promote the progress of science and useful arts”? If so, how?

The Copyright Clause does not permit copyright protection for AI-generated material. *See* responses to Questions 19 and 20. The Copyright Clause speaks of “...securing for limited times to *authors and inventors* the exclusive right to their respective writings and discoveries” (emphasis added).¹⁰⁹ It is difficult to read that clause without concluding that the Constitution was intended to protect humans, not non-humans. All courts that have considered this question agree. *See also* discussion of the Thaler case in response to Question 1.

¹⁰⁹ U.S. Const. art. I, § 8, cl. 8.

Infringement

22. Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right? If so, in what circumstances?

Yes. Exclusive rights of sound recording copyright owners, including the right of reproduction and the right to create a derivative work, are clearly implicated when the AI outputs are identical to pre-existing sound recordings or when the outputs include identifiable portions of pre-existing sound recordings. Even where the naked ear cannot identify full or partial copies of pre-existing sound recordings in the outputs, the outputs can still be infringing to the extent that the model reproduced copyrighted sound recordings “by repressing, transcribing, recapturing off the air, or any other method.”¹¹⁰ See response to Question 2 for a fuller explanation of why Section 114(b), properly construed, is not a bar to claims of infringing audio outputs. In addition, as explained more fully in response to Question 14, an infringement analysis of the AI outputs does not end the inquiry. There may also be infringement of the copyrighted works at the data collection and curation phase, the ingestion phase, and by the AI model itself as an infringing derivative work.

23. Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?

It is too soon to answer this question. It may be that, in light of technological developments involving the reprocessing of sounds, *see* discussion of Section 114(b) in the response to Question 2, traditional notions of how infringement is assessed for sound recordings are not appropriate in the AI context. It will depend on the facts and circumstances of each case.

24. How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used? Are existing civil discovery rules sufficient to address this situation?

Proving the element of copying when the AI developer does not maintain or make available records of which copyrighted works it (and others in the AI supply chain) has used is a thorny problem and highlights why recordkeeping and disclosure requirements are *so* important. Those requirements are not just necessary for transparency as an end goal itself; they are necessary to allow for effective enforcement of rights, which is itself important to maintain the incentives for creating, distributing, and marketing the new sound recordings and other creative works that enrich all our lives.

Existing civil discovery rules may not be sufficient to address this situation, in part because Section 411’s requirement of identifiable, registered works in suit often makes it impossible for an aggrieved rights owner even to get through the courthouse door without knowing what works were infringed in the development of the AI system and its outputs. As a result, a copyright owner may not even to be able to invoke existing discovery rules, regardless of whether those

¹¹⁰ H.R. Rep. No. 94-1476, at 106 (1976).

rules would otherwise be sufficient. This “chicken and egg problem” only underscores the need for transparency and recordkeeping obligations, coupled with a pre-litigation subpoena process with evidentiary presumptions, as outlined in response to Questions 5 and 15.2 above.

25. If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?

Like all questions regarding infringement, the answers will be fact-dependent. That said, everyone who reproduces a copyrighted sound recording, prepares derivative works, distributes, or publicly performs a copyrighted sound recording without authorization is an infringer. Copyright liability is broad enough to encompass multiple infringers (it is joint and several, after all) and, depending on the circumstances, the developer of the model, the developer of the system, and users of the system who create infringing outputs could, under appropriate circumstances, all be directly liable.¹¹¹ Secondary liability also plays a role here, and anyone who satisfies the tests for it (whether inducement, contributory infringement, or vicarious liability) should also be held responsible.

25.1. Do “open-source” AI models raise unique considerations with respect to infringement based on their outputs?

No. Anyone who makes a copy or derivative work is an infringer; it doesn’t matter whether the tool they use is open-source or proprietary. A concern with open-source AI models is that the distributed nature of open-source development may make enforcement of rights more challenging given the investigative difficulties of identifying all of the relevant contributors. Another concern is the proliferation of forked versions of those AI models, which again can make enforcement challenging. Nevertheless, the infringement that occurs in the development and use of such models is no less violative of copyright than with proprietary models.

26. If a generative AI system is trained on copyrighted works containing copyright management information, how does 17 U.S.C. 1202(b) apply to the treatment of that information in outputs of the system?

Section 1202(b) proscribes the removal of copyright management information (CMI). Section 1202(b) applies to the removal or alteration of CMI, including both the ingestion of copyrighted materials that may be stripped of CMI and the generation and distribution of potentially infringing output that contains altered or false CMI (or from which CMI has been removed). If a generative AI system is trained on copyrighted works that were originally distributed by the copyright owner with CMI, and if any downstream user – such as the creator of the AI training dataset or the AI developer who uses the dataset to train the AI system – intentionally removed or altered the CMI in the process, then that user could potentially be liable for a violation of Section

¹¹¹ 5 Melvin Nimmer & David Nimmer, *Nimmer on Copyright* § 14.04[E][2][d][i] (2020) (“When two or more persons have joined in or contributed to a single infringement of a single copyright, each is jointly and severally liable; [and in such] circumstances, in a single infringement action...”) (footnote omitted).

1202(b).¹¹² Stripping out the CMI is not just a “technical violation.” Without the CMI, recordkeeping at scale, which requires some degree of automation, is much more difficult.

27. Please describe any other issues that you believe policymakers should consider with respect to potential copyright liability based on AI-generated output.

Intentionally left blank.

Labeling or Identification

28. Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?

Works generated purely with AI should be labeled and identified as such, as well as any works that are substantially modified with AI to mimic a sound recording artist’s name, image, voice, or likeness without appropriate authorization. This is consistent with the HAC Principles, which state that “content generated solely by AI should be labeled describing all inputs and methodology used to create it – informing consumer choices, and protecting creators and rightsholders.”¹¹³

Such labeling and identification should be tamper-resistant, and could include digital watermarking, metadata identification, and a visual or audio indication to the consumer that the work is AI-generated. Similar to the rationale behind the Integrity, Notification, and Fairness in Online Retail Marketplaces for Consumers Act (the “INFORM Consumers Act”),¹¹⁴ providing such transparency can help deter the use of AI systems to infringe upon a person’s rights, including their rights in their creative expression.

Interest in labeling is widespread. The Administration’s agreement with various technology companies concerning AI issues includes a commitment to watermark or similarly identify AI-generated content, among other things.¹¹⁵ Legislation pending in Congress would make it an unfair trade practice not to put a specified disclaimer on generative AI outputs, with details to be worked out by the FTC.¹¹⁶ Several speakers at the Senate Judiciary Subcommittee on Privacy, Technology and the Law Hearing titled *Oversight of A.I.: Rules for Artificial Intelligence*, May 16, 2023, including Sam Altman, the CEO of OpenAI, agreed that consumers should be alerted

¹¹² See *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2nd Cir. 2020).

¹¹³ Core Principles for Artificial Intelligence Applications in support of human creativity & accomplishment, Principle 6, Human Artistry Campaign, <https://www.humanartistrycampaign.com/> (last visited Oct. 27, 2023).

¹¹⁴ Consolidated Appropriations Act, 2023, Pub. L. 117–328, div. BB, title III, §301, 136 Stat. 5555 (codified at 15 U.S.C. § 45f) (2022) (Took effect on June 27, 2023, which requires online marketplaces to collect, verify, and make available to buyers certain identification information for “high-volume third party sellers” on their platforms).

¹¹⁵ See *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*, The White House (July 21, 2023), <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>.

¹¹⁶ See AI Disclosure Act of 2023, H.R. 3831, 118th Cong., 1st Sess.

when generative AI is used to create video or audio.¹¹⁷ There is also evidence of market interest in labeling. For example, Instagram is reportedly testing labels for posts generated by AI.¹¹⁸ And Adobe Stock requires that uploaded AI-generated images be identified.¹¹⁹

Of course, such labeling and identification, in and of itself, is insufficient to deter the unauthorized use of copyrighted works or the unauthorized use of an artist's name, image, voice or likeness and the associated financial, reputational, or other harms that could result from such conduct. In addition, while failure to label appropriately could and should lead to appropriate penalties, accurate labeling does not absolve liability for any violations of copyright law or the unauthorized use of voice or likeness under existing state or any newly created federal law that may arise from the generative AI.

28.1. Who should be responsible for identifying a work as AI-generated?

Where a work is wholly generated by AI, there are several people in the supply chain who should ensure that the work is appropriately labeled and identified. In the first instance, the AI developer and deployer should be responsible for such labeling and identification. The person or entity that prompted the AI system to generate the work should also be responsible for ensuring that the work is properly labeled and identified, and that such labeling and identification is not tampered with, degraded, or destroyed. To the extent the work is distributed to online intermediaries that make the work accessible to end users, they too should be obligated to make such labeling and identification readily accessible and to not tamper with, degrade, or destroy such labeling and identification.

28.2. Are there technical or practical barriers to labeling or identification requirements?

We presume there are no insurmountable barriers as several AI companies have voluntarily committed to provide such identification, such as through watermarking, through the White House voluntary commitments on AI. However, we have concerns with the use of bespoke watermarking technology by each individual AI company. To the extent there could be some standardization or consistency for watermarking in at least each media type, that would help downstream providers make such watermarks readable and accessible to the public, which is essential to fulfill the goal of informing consumers. We also have concerns with the fragility of

¹¹⁷ See Darrell M. West, *Senate hearing highlights AI harms and need for tougher regulation*, Brookings Institute (May 17, 2023) <https://www.brookings.edu/blog/techtank/2023/05/17/senate-hearing-highlights-ai-harms-and-need-for-tougher-regulation/>. See also *OpenAI CEO Testifies on Artificial Intelligence*, C-SPAN (May 16, 2023) <https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence> (video of the hearing, for example, at 47:56 and 1:18:13. Just as industry best practices (backstopped by enforcement under current authority) have been used to address concerns such as paid reviews from online influencers and sponsored search content, stakeholders should work collaboratively to develop standards for technologies to identify content generated solely by AI).

¹¹⁸ See Sarah Kearns, *Instagram Is Testing Out Labels for Posts Generated by AI*, Hypebeast (Aug. 1, 2023), <https://hypebeast.com/2023/8/instagram-labeling-ai-generated-posts-feature>.

¹¹⁹ See *Generative AI Content*, Adobe (Sept. 26, 2023), <https://helpx.adobe.com/stock/contributor/help/generative-ai-content.html>.

the watermarks that may be used and recommend that they be tested to confirm they survive standard digital file manipulations.¹²⁰

28.3. If a notification or labeling requirement is adopted, what should be the consequences of the failure to label a particular work or the removal of a label?

The consequences of failure to label would depend on the ultimate scope of the labeling requirement, which actor in the chain is at fault, if the fault is intentional or a mere mistake, and the anticipated consequences of such failure to comply. Notwithstanding those nuances, it would make sense – given the goals of a labeling requirement – to use a fraud or deceptive or unfair trade practices framework in determining the appropriate consequences for failure to comply. Once again, whatever the consequences for failure to label, they would not absolve the actors of any consequences they may face if their actions infringe copyright or trademark or violate an artist’s voice or likeness rights.

29. What tools exist or are in development to identify AI-generated material, including by standard-setting bodies? How accurate are these tools? What are their limitations?

Several tools are in development to identify AI-generated material, including the following:

- The Coalition of Content Provenance and Authenticity (C2PA) has been developing technical standards to show the provenance of digital content. This includes methods to determine what AI model was used to generate the work and to validate if the work was produced by an AI model.¹²¹
- DDEX, which develops messaging standards for use in the music industry, is looking at how its standards could be used to communicate information about AI in relation to sound recording content.
- A.I.O.K. is looking to develop a certification standard to identify music that was created with AI in an ethical manner that is respectful of intellectual property rights.¹²²
- Several AI companies agreed to the White House voluntary commitments to identify AI-generated material, such as through the use of digital watermarks. While this area is promising, care needs to be taken to ensure such watermarks are both robust and tamper-resistant and do not degrade when standard manipulations are done to the digital files.

In addition, academics have been studying this issue in connection with voice generally and for singing. Please see section 6 of Annex A for some of these papers.

¹²⁰ See, e.g., Xuandong Zhao, *Invisible Image Watermarks are Provably Removable Using Generative AI*, arXiv, Aug. 6, 2023, available at <https://arxiv.org/pdf/2306.01953.pdf>; see also Kate Knibbs, *Researchers Tested AI Watermarks – and Broke All of Them*, Wired (Oct. 3, 2023), <https://www.wired.com/story/artificial-intelligence-watermarking-issues/>.

¹²¹ See *AI-ML Model Manifest*, under *Guidance for Artificial Intelligence and Machine Learning*, C2PA, https://c2pa.org/specifications/specifications/1.3/ai-ml/ai_ml.html#_ai_ml_model_manifest (last visited Oct. 25, 2023).

¹²² AI:OK, <https://ai-ok.org/> (last visited Oct. 27, 2023).

Finally, we note that some companies are offering or developing services to identify AI-generated material.¹²³

Additional Questions About Issues Related to Copyright

30. What legal rights, if any, currently apply to AI-generated material that features the name or likeness, including vocal likeness, of a particular person?

Currently, legal protection for AI-generated material that features the name or likeness, including vocal likeness, of a particular person is protected by a patchwork of largely state laws as well as the federal Lanham Act (in certain circumstances). As discussed in response to Question 5, we are advocating for a federal right that would provide baseline federal protections against unauthorized uses of artists' voices and likenesses.

31. 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?

Yes, Congress should establish a new federal right, similar to state law rights of publicity (but not preempting those rights), that would apply to AI-generated material. *See* responses to Questions 5 and 30.

32. 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?

To the extent that AI systems are based on, or derive their value from, a particular artist's identity, that artist should be protected by laws governing the use of an individual's brand or identity (such as the individual's voice or likeness), including the Lanham Act and laws regarding rights of publicity and unfair competition. In addition, the artist's privacy rights may be implicated. The federal law for which we are advocating should help to secure these rights nationwide and facilitate effective enforcement. With respect to copyright law, however, we agree with cases like the ones involving Katy Perry, Ed Sheeran, and Led Zeppelin¹²⁴ that rejected infringement claims based only on elements of an artist's style or other uncopyrightable elements.

33. With respect to sound recordings, how does section 114(b) of the Copyright Act relate to state law, such as state right of publicity laws? Does this issue require legislative attention in the context of generative AI?

¹²³ *See* Aminu Abdullahi, *Top 10 AI Detector Tools for 2023*, eWeek (Aug. 4, 2023), www.eweek.com/artificial-intelligence/ai-detector-software/.

¹²⁴ *See* Kathryn Griffin v. Ed Sheeran, case number 1:17-cv-05221 (S.D.N.Y. May 4, 2023) (Ed Sheeran's song "Thinking Out Loud" did not infringe Marvin Gaye's song "Let's Get it On"); Marcus Gray v. Katheryn Hudson, Case No. 20-55401 (9th Cir. Apr. 15, 2020) (Katy Perry's song "Dark Horse" did not infringe Marcus Gray's song "Joyful Noise"); Michael Skidmore v. Led Zeppelin, et. al., 952 F. 3d 1051 (9th Cir. Mar. 9, 2020), cert den'd 141 S.Ct. 453 (Oct. 5, 2020) (Led Zeppelin's song "Stairway to Heaven" did not infringe Spirit's song "Taurus").

Section 114(b) of the Copyright Act and state right of publicity law address different concerns.¹²⁵ Section 114(b) was intended to insulate from potential copyright infringement liability the independent creation of a recording of a new performance that sounds like a preexisting recording, and would otherwise be infringing, because “one performer deliberately sets out to simulate another’s performance as exactly as possible.”¹²⁶ Such an imitation is sometimes referred to as a “soundalike” recording. By contrast, right of publicity law protects recognizable human personas (sometimes including voices), as opposed to particular copyrighted works. Given their different aims, there is substantial room for these two legal regimes to coexist.

Presumably the Office is asking this question in the context of an NOI directed to AI issues because some AI systems have permitted users to generate audio simulating the voice of a particular performing artist without their permission. Promotion of such systems using the artist’s name or likeness, and dissemination of such AI-generated audio simulations, particularly when that is done using the artist’s name or likeness, is an unfair practice that disrespects human creators and human creativity.¹²⁷

We do not believe that such claims are preempted by copyright law. It is clear Congress did not intend copyright law generally to preempt state law concerning the right of publicity and similar state law rights.¹²⁸ That is, Congress seems to have intended to permit the creation and dissemination of soundalike recordings without copyright infringement liability, while simultaneously allowing state law to circumscribe such activities in certain circumstances. Consistent with that intention, the weight of authority involving soundalike recordings holds that state law claims for imitation of a distinctive voice in a soundalike recording are not preempted by Section 114(b).¹²⁹

The only arguable exception of which we are aware is *Romantics v. Activision Publishing, Inc.*¹³⁰ The result in that case, from the Eastern District of Michigan, seems to reflect the confluence of

¹²⁵ See Christopher T. Zirpoli, Cong. Rsch. Serv., LSB11052, Artificial Intelligence Prompts Renewed Consideration of a Federal Right of Publicity 2, (2023), <https://crsreports.congress.gov/product/pdf/LSB/LSB11052> (“[T]he ROP is distinct from the forms of IP already protected by federal law...”).

¹²⁶ H.R. Rep. No. 94-1476 at 106 (1976).

¹²⁷ See, e.g., *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093, 1098-1102 (9th Cir. 1992) (imitation of Tom Waits’ voice); *Midler v. Ford Motor Co.*, 849 F.2d 460, 463-64 (9th Cir. 1988) (imitation of Bette Midler’s voice); *Lahr v. Adell Chemical Co.*, 300 F. 2d 256, 259 (1st Cir. 1962) (imitation of Bert Lahr’s voice).

¹²⁸ See H.R. Rep. No. 94-1476 at 132 (1976) (“The evolving common law rights of ‘privacy,’ ‘publicity,’ and trade secrets, and the general laws of defamation and fraud, would remain unaffected as long as the causes of action contain elements, such as an invasion of personal rights or a breach of trust or confidentiality, that are different in kind from copyright infringement.”).

¹²⁹ See *Waits*, 978 F.2d at 1099-1100; *Midler*, 849 F.2d at 462.

¹³⁰ *Romantics v. Activision Publ’g, Inc.*, 574 F. Supp. 2d 758, 762-3, 766-68 (E.D. Mich. 2008) (involving a new recording of a musical composition originally written and performed by members of the plaintiff performing group that was used in the game *Guitar Hero*. The plaintiffs did not own the copyright to the composition, which was used under a license, and sued for a violation of their rights of publicity (among other things). After deciding against the plaintiffs on state law grounds, the court opined in dicta that even if the plaintiffs had had a claim under state law, that claim would have been preempted by the First Amendment and copyright law. It is not clear that the *Romantics* court’s copyright preemption dictum should be given any weight at all. At most, it suggests a narrow scope of preemption where state law otherwise would in effect allow a performer to control copying of a particular recording).

two factors – first, that the relevant state’s law “does not recognize an analogous claim [for misappropriation of identity] based on the sound of a voice,”¹³¹ and second, that the court did not view the plaintiffs as having a recognizable “sound . . . identifiable separate from the Song.”¹³² Because the court did not believe that the plaintiffs’ persona was recognizably used in the game, it viewed their right of publicity claim as equivalent to a copyright claim relating to a specific copyrighted recording of the one musical composition that was the focus of the action. That is very different from the claims at issue in *Waits* and *Midler*, which involved clearly recognizable imitations of distinctive voices that transcended particular recordings, and which those courts viewed as quite distinct from copyright claims.¹³³ It is not clear that the *Romantics* court’s copyright preemption dictum should be given any weight at all. At most, it suggests a narrow scope of preemption where state law otherwise would in effect allow a performer to control copying of a particular recording.

Such a narrow scope of preemption would be analogous to judicial decisions outside the specific context of soundlike recordings, where courts have tended to find that right of publicity claims are preempted by copyright only when an action is an effort by someone whose persona is embodied in a work, but who is not the copyright owner of the work, to assert the equivalent of copyright claims with respect to use of the work.¹³⁴ More generally, courts have tended to find that right of publicity claims are not preempted by copyright, particularly when a voice or likeness is used for a commercial purpose or there is an implied endorsement by the person depicted.¹³⁵

Case law concerning copyright preemption of rights of publicity is still developing, and we are not aware of any cases addressing that issue in the specific context of AI. Whether courts

¹³¹ *Id.* at 764.

¹³² *Id.*

¹³³ *See* *Midler*, 849 F.2d at 462 (“A voice is not copyrightable.”); *see also* *Waits*, 978 F.2d at 1100.

¹³⁴ *See, e.g.*, *In re Jackson*, 972 F.3d 25, 35-54 (2d Cir. 2020) (right of publicity claim involving use of sample of plaintiff’s recording preempted where gravamen of the claim was use of sample, not use of plaintiff’s identity); *Maloney v. T3 Media, Inc.*, 853 F.3d 1004, 1010-20 (9th Cir. 2017) (right of publicity claim involving noncommercial distribution of photographs of athletes preempted where claim sought to control use of copyrighted image for expressive purposes); *Dryer v. Nat’l Football League*, 814 F. 3d 938, 942-44 (8th Cir. 2016) (football players’ right of publicity claim involving use of footage from games in which they played in documentaries were preempted); *Ray v. ESPN, Inc.*, 783 F.3d 1140, 1142-44 (8th Cir. 2015) (right of publicity claim involving re-telecast of recorded wrestling performances preempted where claim based solely on depiction of plaintiff in copyrighted recordings); *Laws v. Sony Music Entertainment, Inc.*, 448 F.3d 1134, 1139-45 (9th Cir. 2006) (right of publicity claim involving use of licensed sample of plaintiff’s recording preempted); *Fleet v. CBS, Inc.*, 50 Cal. App. 4th 1911, 1918-24 (1996) (actors’ right of publicity claim involving distribution of film in which they appeared were preempted).

¹³⁵ *See, e.g.*, *Facenda v. N.F.L. Films, Inc.*, 542 F.3d 1007, 1026-32 (3d Cir. 2008) (claiming use of clips of famous sports announcer’s voice in promotion of video game product not preempted); *Toney v. L’Oreal U.S.A, Inc.*, 406 F.3d 905, 908-11 (7th Cir. 2005) (claiming use of model’s likeness in advertising not preempted even though a copyrighted photograph of her was used); *Downing v. Abercrombie & Fitch*, 265 F.3d 994, 1003-1005 (9th Cir. 2001) (claiming use of plaintiffs’ likeness in copyrighted photograph not preempted when photograph used in advertising material); *Landham v. Lewis Galoob Toys, Inc.*, 227 F. 3d 619, 623-24 (6th Cir. 2000) (claiming use of actor’s likeness in action figure of character he had played not preempted (although it failed on its merits)); *KNB Enters. v. Matthews*, 92 Cal. Rptr. 2d 713, 719-23 (Ct. App. 2000) (claiming use of model’s likeness in copyrighted photograph not preempted when photograph used for commercial purpose without permission of the copyright owner).

properly recognize the distinction between Section 114(b) and rights of publicity explained above as the law develops with respect to generative AI is an issue that warrants attention. Legislative clarification is not clearly necessary but may prove to be helpful. *See also* Response to Question 2 above that addresses the substance of Section 114(b).

34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study.

Intentionally left blank.

Summary and Conclusion

We thank the Office for raising such thorough, thoughtful, and timely questions concerning the application of existing copyright law to generative AI, the potential need for legislative or regulatory solutions to the challenges posed by AI, and for including in its study important copyright-adjacent issues such as creators' rights to their voice and likeness.

A2IM and RIAA appreciate the opportunity to comment on these important issues that are already affecting our member companies, the artists whose sound recordings they reproduce, distribute, market, and promote, and the fans who listen to the music. We conclude with this quote from the HAC: "Only humans are capable of communicating the endless intricacies, nuances, and complications of the human condition through art - whether it be music, performance, writing, or any other form of creativity. Developments in artificial intelligence are exciting and could advance the world farther than we ever thought possible. But AI can never replace human expression and artistry. As new technologies emerge and enter such central aspects of our existence, it must be done responsibly and with respect for the irreplaceable artists, performers, and creatives who have shaped our history and will chart the next chapters of human experience."

Respectfully submitted,



Dr. Richard James Burgess MBE
President and CEO
A2IM (The American Association of Independent Music)
132 Delancey Street
2nd Floor
New York, NY 10002
richard@a2im.org

[Signatures continue on next page]

A handwritten signature in cursive script that reads "Susan Chertkof".

Susan B. Chertkof
Senior Vice President, Business & Legal Affairs
Recording Industry Association of America, Inc.
1000 F Street, N.W.
Floor 2
Washington, D.C. 20004
(v) 202-775-0101
(f) 202-775-7253
schertkof@riaa.com

Annex A – List of Papers Relevant to this NOI

1. Identification of Training Data:

Nicholas Carlini et al., *Extracting Training Data from Diffusion Models*, arXiv.org, Jan. 30, 2023, available at <https://arxiv.org/pdf/2301.13188.pdf>. The authors of this paper find that “diffusion models are much less private than prior generative models such as GANs, and that mitigating these vulnerabilities may require new advances in privacy-preserving training.”

Nicholas Carlini et al., *The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks*, arXiv.org, July 16, 2019, available at <https://arxiv.org/abs/1802.08232v3>. The authors of this paper “show that unintended memorization is a persistent, hard to avoid issue that can have serious consequences.”

Yizhan Huang et al., *Do Not Give Away My Secrets: Uncovering the Privacy Issue of Neural Code Completion Tools*, arXiv.org, Sept. 14, 2023, available at <https://arxiv.org/pdf/2309.07639.pdf>. The authors of this paper extracted 2,702 hard-coded credentials from Copilot and 129 secrets from CodeWhisperer, which they state “raise severe privacy concerns of the potential leakage of hard-coded credentials in the training of” neural code completion tools.

Yan Pang et al., *White-Box Membership Interference Attacks Against Diffusion Models*, arXiv.org, Aug. 11, 2023, available at <https://arxiv.org/pdf/2308.06405.pdf>. This article provides an analysis of existing membership interference attacks on diffusion models and offers another approach for such attacks.

Gowthami Somepalli et al., *Diffusion Art of Digital Forgery? Investigating Data Replication in Diffusion Models*, arXiv.org, Dec. 12, 2022, available at <https://arxiv.org/pdf/2212.03860.pdf>. The authors of this paper “identify cases where diffusion models, including the popular Stable Diffusion model, blatantly copy from their training data.”

Ruixiang Tang et al., *Did you Train on My Dataset? Towards Public Dataset Protection with Clean-Label Backdoor Watermarking*, arXiv.org, Apr. 10, 2023, available at <https://arxiv.org/pdf/2303.11470.pdf>. The authors of this paper propose a watermarking framework that uses imperceptible perturbations to mark datasets to be able to identify their unauthorized use in AI models. They found that their “experiments on text, image and audio datasets demonstrate that the proposed framework effectively safeguards datasets with minimal impact on original task performance.”

2. Influence of Training Data on Output

Roger Grosse, et al, *Studying Large Language Model Generalization with Influence Functions*, arXiv.org, Aug. 7, 2023, available at <https://arxiv.org/pdf/2308.03296.pdf>. In this article, the authors observed that for famous passages, “invariably, the top influential sequences returned by our scan contained the exact famous passages” and that “overlaps between the influence query and the scanned sequences do in fact lead to high influence scores and that our influence scans are able to find matches, at least for clear-cut cases of memorization.”

Zayd Hammoudeh et al., *Training Data Influence Analysis and Estimation: A Survey*, arXiv.org, June 27, 2023, available at <https://arxiv.org/pdf/2212.04612.pdf>. This article “reviews numerous methods with different perspectives on – and even definitions of – training data influence.”

Garima Pruthi et al., *Estimating Training Data Influence by Tracing Gradient Descent*, arXiv.org, Nov. 14, 2020, available at <https://arxiv.org/pdf/2002.08484.pdf>. The authors state that this approach “applies

to any machine learning model trained using stochastic gradient descent or a variant of it, agnostic of architecture, domain and task.”

3. Amount of Training Data Required

Wen-Chin Huang et al., *AAS-VC : On the Generalization Ability of Automatic Alignment Search Based on Non-Autoregressive Sequence to Sequence voice Conversion*, arXiv.org, Sept. 15, 2023, available at <https://arxiv.org/pdf/2309.07598.pdf>. The authors of this paper show experimental results that “AAC-VC can generalize better to a training dataset of only 5 minutes.”

Shuo Yang et al, *Dataset Pruning: Reducing Training Data by Examining Generalization Influence*, arXiv.org, Feb. 27, 2023, available at <https://arxiv.org/pdf/2205.09329.pdf>. In this paper, the authors propose an optimization-based dataset pruning method to remove redundant training examples with minor impact on the model’s performance. Their proposed method pruned “40% training examples on the CIFAR-10 dataset,” having the convergence time with only “1.3% test accuracy decrease.”

4. Unlearning of Training Data

Than Tam Nguyen et al., *A Survey of Machine Unlearning*, arXiv.org, Oct. 21, 2022, available at <https://arxiv.org/pdf/2209.02299v5.pdf>. This paper seeks to provide an introduction to machine unlearning and its formulations, design criteria, removal requests, algorithms, and applications.

Jie Xu et al., *Machine Unlearning: Solutions and Challenges*, arXiv.org, Aug. 14, 2023, available at <https://arxiv.org/pdf/2308.07061.pdf>. According to its authors, this paper provides “an in-depth critical analysis of existing machine unlearning techniques, highlighting their strengths, limitations, and challenges.”

Xulong Zhang et al., *Machine Unlearning Methodology base on Stochastic Teacher Network*, arXiv.org, Aug. 28, 2023, available at <https://arxiv.org/pdf/2308.14322.pdf>. In this paper, the authors propose “using a stochastic network as a teacher to expedite the mitigation of the influence caused by forgotten data on the model.”

Ronen Eldan and Mark Russinovich, *Who’s Harry Potter? Approximate Unlearning in LLMs*, arXiv.org, Oct. 4, 2023, available at <https://browse.arxiv.org/pdf/2310.02238.pdf>. In this paper, the authors acknowledge that large language models (LLMs) “are trained on massive internet corpora that often contains copyright infringing content” and they propose a “novel technique for unlearning a subset of the training data from an LLM, without having to retrain it from scratch.”

5. Music Generation Models and Concerns with Copyright Infringement

Peling Lu et al., *MuseCoco: Generating Symbolic Music from Text*, arXiv.org, May 31, 2023, available at <https://arxiv.org/pdf/2306.00110.pdf>. In this paper, the authors caution that “utilization of generative AI for creative purposes often raises concerns regarding copyright and ownership, which necessitates careful consideration going forward.”

Giorgio Mariani et al., *Multi-Source Diffusion Models for Simultaneous Music Generation and Separation*, arXiv.org, Feb. 9, 2023, available at <https://arxiv.org/pdf/2302.02257.pdf>. In discussing their model to generate music and provide source separation, the authors note that they trained their model with creative commons licensed data, and that “[t]raining our model on copyrighted audio data can result in data forgery” and they “strongly discourage such behavior.”

Andrea Angostinelli et al., *MusicLM : Generating Music from Text*, arXiv.org, Jan. 26, 2023, available at <https://arxiv.org/pdf/2301.11325.pdf>. In this paper, the authors acknowledge “the risk of potential misappropriation of creative content associated to the use-case” and “strongly emphasize the need for more future work in tackling these risks associated to music generation,” noting they have “no plans to release models at this point.”

Julia Barnett, *The Ethical Implications of Generative Audio Models: A Systematic Literature Review*, arXiv.org, July 7, 2023, available at <https://arxiv.org/pdf/2307.05527.pdf>. In this article, the author quantifies the lack of ethical consideration in generative audio research, and identifies key areas of potential harm, including fraud, deep-fakes, and copyright infringement.

Katherine Lee et al., *Talkin’ Bout AI Generation: Copyright and The Generative-AI Supply Chain*, arXiv.org, Sept. 14, 2023, available at <https://arxiv.org/ftp/arxiv/papers/2309/2309.08133.pdf>.

Luca Lazendorfer et al., *DISCO-10M: A Large-Scale Music Dataset*, arXiv.org, June 23, 2023, available at <https://arxiv.org/pdf/2306.13512.pdf>. In this paper, the authors acknowledge that “[c]opyright and licensing agreements are a complex issue, particularly when it comes to big data collection for training large machine learning models,” and they “acknowledge the concerns of artists regarding the potential negative impact on their artistic work.”

Jade Copet et al, *Simple and Controllable Music Generation*, arXiv.org, June 8, 2023, available at <https://arxiv.org/pdf/2306.05284.pdf>. The authors here note that their model was trained on 20K hours of licensed music data.

6. Deep Fake Voice/Singing Detection

Jordan Bird et al, *Real-time Detection of AI-Generated Speech for DeepFake Voice Conversion*, arXiv.org, Aug. 24, 2023, available at <https://arxiv.org/pdf/2308.12734.pdf>.

Yongzi Zang et al, *SingFake: Singing Voice Deepfake Detection*, arXiv.org, Sept. 14, 2023, available at <https://arxiv.org/pdf/2309.07525.pdf>.

7. Problems with Training an AI System with AI Generated Data

Ilia Shumailov et al, *The Curse of Recursion: Training on Generated Data Makes Models Forget*, arXiv.org, May 31, 2023, available at <https://arxiv.org/pdf/2305.17493v2.pdf>. The authors of this article find that “the value of data collected about human interactions with systems will be increasing valuable in the presence of content generated by LLMs in data crawled from the Internet.”

Sina Alemohammad et al, *Self-Consuming Generative Models go MAD*, arxiv.org, July 4, 2023, available at <https://arxiv.org/pdf/2307.01850.pdf>. The authors of this article find that across all scenarios, “without enough fresh real data in each generation of an autophagous loop, future generative models are doomed to have their quality (precision) or diversity (recall) progressively decrease.”

8. Watermarks

Xuandong Zhao, *Invisible Image Watermarks are Provably Removable using Generative AI*, arXiv.org, Aug. 6, 2023, available at <https://arxiv.org/pdf/2306.01953.pdf>.

ANNEX B
SAMPLE LEGISLATIVE LANGUAGE FOR A NEW ADMINSTRATIVE SUBPOENA

SEC X. SUBPOENA FOR RECORDS AND COPIES

- (1) Request.— An intellectual property rights owner or a person authorized to act on such owner’s behalf may request the clerk of any United States district court to issue a subpoena to a model developer or deployer for disclosure of the records related to, and copies of the data used to train, the generative AI model.
- (2) Content of Request.— The request may be made by filing with the clerk—
 - a. A proposed subpoena; and
 - b. A sworn declaration to the effect that (i) the intellectual property rights owner or person authorized to act on such owner’s behalf has a subjective good faith belief that the model developer or deployer trained the generative artificial intelligence model in violation of such owner’s rights or otherwise used some or all of such owner’s intellectual property to train the generative artificial intelligence model, (ii) the purpose for which the subpoena is sought is to obtain the records or copies of the data used to train the generative AI training model to determine whether the model developer or deployer has violated such owner’s rights in connection with the generative artificial intelligence model, and (iii) such records and copies will only be used for the purpose of protecting such owner’s rights under state or federal law.
- (3) Contents of Subpoena.— The subpoena shall authorize and order the model developer or deployer receiving the notification and the subpoena to expeditiously disclose to the owner or person authorized by the owner all such records and copies.
- (4) Basis for Granting Subpoena.— If the proposed subpoena is in proper form, and the accompanying declaration is properly executed, the clerk shall expeditiously issue and sign the proposed subpoena and return it to the requester for delivery to the model developer or deployer.
- (5) Actions of Model Developer or Deployer Receiving Subpoena.— Upon receipt of the issued subpoena, the model developer or deployer shall expeditiously disclose to the owner or person authorized by such owner the records and copies required by the subpoena.
- (6) Rules Applicable to Subpoena.— Unless otherwise provided by this section or by applicable rules of the court, the procedure for issuance and delivery of the subpoena, and the remedies for noncompliance with the subpoena, shall be governed to the greatest extent practicable by those provisions of the Federal Rules of Civil Procedure governing the issuance, service, and enforcement of a subpoena duces tecum.
- (7) Rebuttable Presumption.— If a model developer or deployer fails to comply with a subpoena issued under this Act or fails to maintain the records and copies as provided in Section 3, such failure shall provide a rebuttable presumption that the model developer or deployer willfully infringed the owner’s intellectual property rights, including, without limitation, the owner’s right of reproduction of copyrighted works under Title 17 U.S.C. § 106(1).

(8) Tolling of Limitations Period. — In addition to any tolling that may be appropriate from the model developer’s or deployer’s lack of transparency regarding the use of copyrighted works to train the applicable generative AI model, the making of a request for a subpoena under Section (1) hereof shall further toll the running of any applicable statute of limitations until such time as the subject model developer’s or deployer’s compliance or failure to comply with the subpoena has been established.

ANNEX C

Comments of

American Association of Independent Music,
American Federation of Musicians,
American Photographic Artists,
American Society of Composers, Authors and Publishers,
Artist Rights Alliance,
Association of American Literary Agents,
Association of Independent Music Publishers,
Black Music Action Coalition,
Broadcast Music, Inc.,
Christian Music Trade Association,
Department for Professional Employees, AFL-CIO,
Folk Alliance International,
Gospel Music Association,
Graphic Artists Guild,
Major League Baseball Players Association,
Major League Soccer Players Association,
Music Artists Coalition,
Nashville Songwriters Association International,
National Football League Players Association,
National Hockey League Players' Association,
National Music Publishers' Association,
Recording Industry Association of America,
Screen Actors Guild-American Federation of Television and Radio Artists,
SESAC Music Group, and
Songwriters of North America,

on the Office of Science and Technology Policy
Request for Information on
National Priorities for Artificial Intelligence
Delivered via regulations.gov

July 7, 2023

The organizations set forth above, described on Appendix A hereto, (“We” or the “Creative Community”), welcome this opportunity to provide comments to the Office of Science and Technology Policy (“OSTP”) on its request for information on the national priorities for artificial intelligence (“RFC”).

1. Introduction

Who We Are. We represent a wide swath of artists, performers, creators, and rights holders who contribute to, define, and promote America’s culture, values, and our national identity. Collectively, our members include actors, artists, athletes, authors, musicians, photographers, songwriters, sound recording artists, writers, and those individuals and American businesses, large and small, that support them. Human creative expression is at the core of what our members do and support, and it is vital for our nation’s culture and economy. We are key stakeholders in AI policy development and should have a seat at the table in any conversations regarding legislation, regulation, or government priorities regarding AI that would impact our creativity and the way it affects our industry and livelihoods.

Our contribution to U.S. Jobs and the U.S. Economy. Our community contributes significantly to U.S. jobs, the U.S. economy, U.S. exports, and our digital economy. In 2021, the total copyright industries employed nearly 16.1 million workers, accounting for 8.14% of all U.S. employment.¹ Those jobs paid an average of \$103,752 in annual compensation, exceeding the U.S. average annual wage by around 29%.² The annual compensation paid to core copyright workers that same year was even higher, amounting to a 51% compensation premium over the average U.S. annual wage.³ In addition, in 2021, the total copyright industries added more than \$2.9 trillion to the GDP, accounting for 12.52% of the U.S. economy.⁴ That same year, sales of select U.S. copyrighted products in overseas markets amounted to \$230.3 billion.⁵ In terms of our digital economy, the total copyright industries accounted for 52.26% of the U.S. digital economy, contributing over 58% to the U.S. digital economy employment.⁶

Opportunities and Threats our Community Faces from Artificial Intelligence. As with other technological developments, the creative community lives at the forefront of, and is building and inspiring, evolutions in artificial intelligence (“AI”) technology. AI already serves as a tool to assist the creative process, allowing for a wider range of people to express themselves creatively. AI also has many valuable uses outside of the creative process itself, including those that amplify fan connections, hone personalized recommendations, identify content quickly and accurately, assist with scheduling, and more. We embrace these technological advances.

At the same time, the unethical development and deployment of AI systems threatens our jobs, livelihoods, identity, and dignity.⁷ For example, artists filed a class action lawsuit against Midjourney and Stability AI claiming that these AI companies infringed the rights of millions of artists by copying web-scraped images of the artists’ works without authorization in order to train Midjourney’s and

¹ Stoner, Robert et al., “IIPA, Copyright Industries in the U.S. Economy, 2022 Report,” Secretariat Economists, prepared for the International Intellectual Property Alliance, Dec. 2022, p. 8, available at https://www.iipa.org/files/uploads/2022/12/IIPA-Report-2022_Interactive_12-12-2022-1.pdf.

² Id.

³ Id.

⁴ Id.

⁵ Id.

⁶ Id.

⁷ The White House President’s Council of Advisors on Science and Technology (PCAST) stated “generative AI models can also be used for malicious purposes, such as . . . impersonating individuals” and “generative AI systems can violate privacy and undermine intellectual property rights.” See “PCAST Working Group on Generative AI Invites Public Input,” May 13, 2023, available at <https://www.whitehouse.gov/pcast/briefing-room/2023/05/13/pcast-working-group-on-generative-ai-invites-public-input/#:~:text=The%20President's%20Council%20of%20Advisors,equitably%2C%20responsibly%2C%20and%20safely%20as>

Stability AI's AI systems.⁸ Similarly, Getty Images, a leading supplier of digital images, sued Stability AI for copying and ingesting millions of Getty Images' copyrighted images without permission to train Stability AI's Stability Diffusion AI system.⁹ Recently, two lawsuits were filed against OpenAI alleging copyright and privacy violations concerning ChatGPT.¹⁰ In addition, academics have identified cases where AI diffusion models produce images that are copied from their training data,¹¹ as well noted concerns with ChatGPT replicating text from copyrighted works in its outputs.¹² Sound recording artists and music labels are confronting an explosion of unauthorized AI-cloned vocals that infringe not only the rights of the artists whose voices are being cloned but also the rights of those that own the musical compositions and sound recordings in each underlying musical track.¹³ These and other related harms will only worsen as AI models are further refined and released to the public.¹⁴

Core Principles. The Administration has made it clear that AI development should be done in a manner that is ethical, lawful, trustworthy, safe, and protects human rights. According to the White House, "[t]he important progress [of AI automated systems] must not come at the price of civil rights or democratic values."¹⁵ Such respect for civil rights and democratic values necessarily includes respecting

⁸ See, e.g., "Generative Artificial Intelligence and the Law," Congressional Research Service, May 11, 2023, available at <https://crsreports.congress.gov/product/pdf/LSB/LSB10922>.

⁹ Id.

¹⁰ Southern, Matt, "ChatGPT Creator Faces Multiple Lawsuits Over Copyright & Privacy Violations Search Engine Journal, July 3, 2023, available at <https://www.searchenginejournal.com/chatgpt-creator-faces-multiple-lawsuits-over-copyright-privacy-violations/490686/>.

¹¹ See Somepalli, Gowthami et al., "Diffusion Art of Digital Forgery? Investing Data Replication in Diffusion Models," Dec. 12, 2022, available at <https://arxiv.org/pdf/2212.03860.pdf> ("We identify cases where diffusion models, including the popular Stable Diffusion model, blatantly copy from their training data."). See also Carlini, Nicholas et al., "Extracting Training Data from Diffusion Models," Jan. 30, 2023, available at <https://arxiv.org/pdf/2301.13188.pdf> ("We show that diffusion models memorize individual images from their training data and emit them at generation time.").

¹² Stokel-Walker, Chris, "ChatGPT seems to be trained on copyrighted books like Harry Potter," New Scientist, May 5, 2023, available at <https://www.newscientist.com/article/2372140-chatgpt-seems-to-be-trained-on-copyrighted-books-like-harry-potter/>; Chang, Kent et al., "Speak, Memory: An Archeology of Books Known to ChatGPT/GPT-4," Apr. 28, 2023, available at <https://arxiv.org/pdf/2305.00118.pdf> ("We find that OpenAI models have memorized a wide collection of copyrighted materials, and that degree of memorization is tied to frequency with which passages of those books appear on the web.").

¹³ See, e.g., Di Placido, Dani, "Thanks to AI, Fake Kanye and Drake Songs are Going Viral on TikTok," Apr. 24, 2023, available at <https://www.forbes.com/sites/danidiplacido/2023/04/24/ai-generated-songs-that-sound-like-kanye-and-drake-are-going-viral-on-tiktok/?sh=1f9bfcf13531>.

¹⁴ It bears observing that many of the leading generative AI companies are either recipients of enormous investments from dominant internet platform companies or themselves affiliated with such companies, raising the risk that the same competitive harms visible today in online search, social media, and user-generated content platforms will repeat themselves as AI services are deployed.

¹⁵ See The White House, Office of Science and Technology Policy, "Blueprint for an AI Bill of Rights," released Oct. 4, 2022, available at <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>. See also the statement from the National Telecommunications and Information Administration ("NTIA") that it desires to develop AI accountability policies that can demonstrate "that AI systems are legal, effective, safe and otherwise trustworthy." 88 Fed. Reg. 22433 (Apr. 13, 2023), and the statement from the International Trade Administration that the "United States seeks to promote the development of innovative and trustworthy AI systems that respect human rights, democratic values, and are designed to enhance privacy protections." 87 Fed. Reg. 50288 (Aug. 16, 2022).

the rights of creators, performers, and other rightsholders in their creations, identities, and dignity.¹⁶

The G7 has espoused similar views. The digital and technology ministers of the G7 countries declared that the G7 “reaffirm their commitment to promote human-centric and trustworthy AI based on the OECD AI Principles,” “oppose the misuse and abuse of AI to . . . threaten the enjoyment of human rights,” and “plan to convene further G7 discussions on generative AI which could include topics such as governance, how to safeguard intellectual property rights including copyright, promote transparency, address disinformation, including foreign information manipulation, and how to responsibly utilize these technologies.”¹⁷ The G7 leaders subsequently confirmed their commitment to these discussions.¹⁸

We agree that the development and deployment of AI systems should be done responsibly, ethically, and with respect for the artists, athletes, creators, and performers who have shaped our history and will chart the next chapters of human experience. At its core, we believe that the approach to responsible AI innovation and deployment requires a human-centric approach.

That is why the signatories to these comments have also signed onto the Human Artistry Campaign.¹⁹ The Human Artistry Campaign calls for policy makers, AI developers, and those that deploy AI to take into account the following principles:

- (viii) technology has long empowered human expression, and AI will be no different;
- (ix) human created works will continue to play an essential role in our lives;
- (x) use of copyrighted works and the use of voices and likenesses of professional performers requires authorization and free-market licensing from all rights holders;
- (xi) governments should not create new copyright or other IP exemptions that allow AI developers to exploit creations without permission or compensation;
- (xii) copyright should only protect the unique value of human intellectual creativity;
- (xiii) trustworthiness and transparency are essential to the success of AI and protection of creators; and
- (xiv) creators’ interests must be represented in policy making.²⁰

With these principles in mind, we offer the following comments:

2. Comments

Protecting rights, safety and national security

*Recordkeeping, Audits, Transparency, and Labeling*²¹ To protect people’s rights and safety, including their

¹⁶ See, e.g., the U.N. Universal Declaration of Human Rights, art. 12 (“No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation.”), art. 27, § 2 (“Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author”), available at <https://www.un.org/en/about-us/universal-declaration-of-human-rights>.

¹⁷ See G7 2023 Hiroshima Summit, “Ministerial Declaration the G7 Digital and Tech Ministers’ Meeting, 30 April 2023,” paragraphs 42 and 47, available at https://g7digital-tech-2023.go.jp/topics/pdf/pdf_20230430/ministerial_declaration_dtm.pdf.

¹⁸ See G7 Hiroshima Leaders’ Communique, para. 38, May 20, 2023.

¹⁹ See <https://www.humanartistrycampaign.com/>.

²⁰ Id.

²¹ Responsive to questions 1 and 3.

intellectual property rights, rights of publicity, rights of privacy, and other human rights, AI systems must be designed and deployed in a manner that is accountable, transparent, and trustworthy. To support this objective, U.S. policy should require AI developers and deployers to keep proper records, ensure that those records are periodically audited, disclose appropriate information to the public about their AI systems, and label content solely generated by their AI systems.²²

Proper record-keeping should include complete documentation about (i) the articulated purpose of the AI model itself and its intended outputs, (ii) the AI system’s overall functioning, (iii) what materials were ingested to develop the AI system (or to fine tune or adapt a pretrained AI system) and in what manner, (iv) the provenance of such materials, including whether any licenses or authorizations were sought or obtained to authorize such use and copies of those licenses or authorizations, (v) the articulated rationale for selecting and using the materials ingested for the AI system’s development, (vi) the individual or organization responsible for the AI system (including who is responsible for ingesting the materials, who is responsible for any foundational AI model, who is responsible for any fine tuning of the AI model, who is deploying the AI system, etc.), (vii) risk assessments concerning the potential misuse and abuse of such a model, and (viii) what parameters and processes were used, and what decisions were made, during the AI system development and deployment. According to NIST, having such record-keeping and documentation concerning risk assessments and processes can “enhance transparency, improve human review processes, and bolster accountability” in AI systems.²³

Such record-keeping should occur at various times within the AI development and deployment process, including when an AI system is developed, when it is fine-tuned or adapted for a particular purpose or use case, when the system is deployed, and when problems with the system are uncovered and analyzed.²⁴ Consistent with generally accepted record-keeping rules in other contexts,²⁵ these records should be kept throughout the AI system’s development and deployment lifecycle, and for a period of at least seven years following its discontinuance. Also, the keepers of such records should be bound to produce them when it is determined to be appropriate by an applicable government agency or court.

The AI systems, and their records, should be audited periodically throughout the AI development and deployment lifecycle.²⁶ These audits should include, among other things, checks on the provenance of

²² As noted by the National Institute of Standards and Technology (NIST), “[m]aintaining the provenance of training data and supporting attribution of the AI system’s decisions to subsets of training data can assist with both transparency and accountability.” National Institute of Standard and Technology, “Artificial Intelligence Risk Management Framework (AI RMF 1.0),” NIST AI 100-1, Jan. 2023, p. 16, available at <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

²³ Id. at p. 22.

²⁴ As Ellen Goodman from GMF explained, these records are necessary “to create relevant audit trails” and such trails “could cover all steps of the AI development process, from the institutional work of problem and purpose definition leading up the initial creation of a system, to the training and development of the system, all the way to retrospective accident analysis.” Goodman, Ellen et al., “AI Audit-Washing and Accountability,” G.M.F., Nov. 15, 2022, p. 18, available at <https://www.gmfus.org/news/ai-audit-washing-and-accountability>.

²⁵ See, e.g., 17 C.F.R. Part 210 (SEC Retention of Records Relevant to Audits and Reviews).

²⁶ See Written Testimony of Professor Gary Marcus before the Senate Judiciary Subcommittee on Privacy, Technology and the Law titled *Oversight of A.I.: Rules for Artificial Intelligence*, May 16, 2023, available at <https://www.judiciary.senate.gov/imo/media/doc/2023-05-16%20-%20Testimony%20-%20Marcus.pdf>. See also the comments from researchers from the Stanford Center for Research on Foundation Models, part of the Stanford Institute for Human Centered Artificial Intelligence and Princeton University’s Center for Information Technology Policy, in response to NTIA’s request for comments on AI Accountability Policy, available at

data sources (including checks for intellectual property implications), AI validation checks that include checks for information leakage (which could result in a security breach, persona data breach, or infringement of copyright), checks on transparency and human oversight during deployment, and checks on long term consequences of AI deployment, including its social impact and model drift.²⁷

For transparency, AI developers and deployers should disclose to the public the purpose of the AI system and its overall functionality, who is the individual or entity responsible for the AI system and their location and contact information, the provenance of the materials ingested during the AI system's development, and basic information to provide algorithmic transparency.²⁸ Similar to the rationale behind the INFORM Consumers Act, providing such transparency can help deter the use of AI systems to infringe upon a person's rights, including their rights in their creative expression and in their identity. The U.S. should also consider providing a mechanism for those claiming to be harmed by the AI system to obtain or be able to inspect applicable records and audits of the AI system to determine whether a person's rights have been infringed or violated. The U.S. should also require that public audits of certain high risk, foundational, or widely used AI systems be made available to the public for inspection.²⁹

Moreover, content generated solely with AI should be labeled as such. Such labeling should also either describe all inputs and methodology used to create that content or include a link or other mechanism that permits the public to access such information. This will help inform consumer choices and protect creators and rights holders. Several speakers at the Senate Judiciary Subcommittee on Privacy, Technology and the Law Hearing titled *Oversight of A.I.: Rules for Artificial Intelligence*, May 16, 2023, including Sam Altman, the CEO of OpenAI, agreed that consumers should be alerted when generative AI is used to create video or audio.³⁰

To the extent technical standards are needed or desired to implement requirements for record-keeping, auditing, or transparency, all of the relevant stakeholders, including creators and copyright holders, should be at the table for the development of such standards. Such standards should not be left solely to AI developers.

<https://hai.stanford.edu/sites/default/files/2023-06/Response-to-Request.pdf> ("AI Researcher Comments to NTIA").

²⁷ See Van Otterloo, "A checklist for auditing AI systems," ICT Institute, Dec. 18, 2022, available at <https://ictinstitute.nl/a-checklist-for-auditing-ai-systems/>.

²⁸ This type of transparency is consistent with the White House Blueprint for an AI Bill of Rights, available at <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>, and Congressional calls for tougher disclosure requirements for AI systems, such as those expressed at the Senate Judiciary Subcommittee on Privacy, Technology and the Law Hearing titled *Oversight of A.I.: Rules for Artificial Intelligence*, May 16, 2023 (a video of which is available at <https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence>) and the House Judiciary IP Subcommittee hearing titled *Artificial Intelligence and Intellectual Property: Part I: Interoperability of AI and Copyright Law*, May 17, 2023 (a video of which is available at https://www.youtube.com/watch?v=Mm1NQ_Kqumw).

²⁹ Among others, the AI Researcher Comments to NTIA recommend this approach.

³⁰ See "Senate hearing highlights AI Harms and need for tougher regulation," Brookings Institute, May 17, 2023, available at <https://www.brookings.edu/blog/techtank/2023/05/17/senate-hearing-highlights-ai-harms-and-need-for-tougher-regulation/>. See also a video of the hearing, available at <https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence>, for example, at 47:56 and 1:18:13. Just as industry best practices (backstopped by enforcement under current authority) have been used to address concerns such as paid reviews from online influencers and sponsored search content, stakeholders should work collaboratively to develop standards for technologies to identify content generated solely by AI.

In the interest of increasing transparency and mitigating risk inherent in AI systems, some have called for the implementation of a regulatory framework that requires registration of some or all AI systems before they may be made available for use.³¹ If the U.S. adopts such a framework, it should include not only registration requirements, but also the record-keeping, auditing, and transparency obligations described above. It should also permit private parties to bring a cause of action against the AI developer or deployer in the event the developer or deployer fails to comply with those obligations. In addition to regulating other high risk, foundational, or highly capable AI systems, such a regulatory framework should also apply to AI systems that mimic an individual's voice, image, or identity, or provide the capability to clone an individual's voice, image, or identity without the consent of the individual.

*Authorization and Licensing.*³² We are especially concerned about AI developers and deployers taking our members' copyrighted works or their name, image, likeness, or voice, without their authorization or consent. As noted above, this is happening not only with large language models and other generative AI systems that have ingested our members' creative expression and output content that directly competes with our members' works, but also with vocal cloning models that let others create synthetic, unauthorized speech or music that mimics artists' voices without their consent or authorization.

The risks of these harms, and ways to mitigate them, have been raised by various U.S. agencies, including NIST in its AI Risk Management Framework and OSTP in the National Blueprint for an AI Bill of Rights, as well as by various members of Congress and others in the AI industry. For example, the NIST AI Risk Management Framework provides that "[t]raining data may also be subject to copyright" and therefore AI developers "should follow applicable intellectual property rights laws."³³ The AI Bill of rights provides that AI designers, developers, and deployers should seek an individual's permission and respect an individual's decisions regarding collection, use, access, transfer, and deletion of a person's data, presumably including biometric and other data about a person's voice or image.³⁴ The Government Accountability Office has noted economic issues from generative AI systems trained on copyrighted, proprietary, or sensitive data, without the owner's or subject's knowledge.³⁵ According to the Select Committee on Artificial Intelligence of the National Science and Technology Council, "it is key that U.S. agencies develop and establish appropriately rigorous standards, policies, and procedures for data sharing, data privacy, and the protection of intellectual property to safeguard data, privacy, and national security" in connection with the development of global AI systems, standards, frameworks.³⁶

³¹ See, e.g., Written Testimony of Sam Altman, Chief Executive Officer, Open AI, Before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, & the Law, May 16, 2023, available at <https://www.judiciary.senate.gov/imo/media/doc/2023-05-16%20-%20Bio%20&%20Testimony%20-%20Altman.pdf>; Microsoft, "Governing AI: A Blueprint for the Future," May 25, 2023, p. 20-22, available at <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw>; and the AI Act bill being considered in the European Union, the May 9, 2023 draft compromise agreement of which is available at <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>.

³² Responsive to question 2.

³³ AI RMF 1.0 at p. 16.

³⁴ See Blueprint for an AI Bill of Rights, available at <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>.

³⁵ See "Science and Tech Spotlight: Generative AI," Jun. 13, 2013, GAO-23-106782, available at <https://www.gao.gov/products/gao-23-106782>.

³⁶ National Artificial Intelligence Research and Development Strategic Plan 2023 Update, a Report by the Select Committee on Artificial Intelligence of the National Science and Technology Council, May 2023, p. 36, available at <https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf>.

Similarly, in hearings held in 2022 and 2023, Representatives Nadler and Ross and Senators Blackburn and Klobuchar all noted the importance of protecting intellectual property and raised concerns with AI systems taking creators' works to train AI systems without authorization from or compensation to the creators.³⁷ On June 21, 2023, Senator Schumer released his SAFE Innovation Framework for bipartisan AI legislation, which, among other things, calls for an accountability objective that will "support our creators by addressing copyright concerns, protect intellectual property, and address liability."³⁸

Within the AI industry, the CEO of Open AI, Sam Altman, stated that "creators deserve control over how their creations are used and what happens beyond the point of them releasing it into the world," and stated further that "the right thing is to make sure [creators] get significant upside benefit" from AI technologies, and "content owners, likenesses, people totally deserve control over how that is used and to benefit from it."³⁹ AI auditing firms call for intellectual property checks to be part of any AI audit.⁴⁰ Professor Gary Marcus, a leading voice on artificial intelligence,⁴¹ raised concerns that AI "systems can easily recreate past works without attribution or credit to the original artists"⁴² and suggests that "our focus should be on figuring on how to build AI that can represent and reason about *values*, rather than simply perpetuating past data."⁴³ Timnit Gebru, founder and executive director of the Distributed AI Research Institute,⁴⁴ quoted with approval the statement that "AI-art generators are trained on enormous datasets, containing millions upon millions of copyrighted images, harvested without their creator's knowledge, let alone compensation or consent. This is effectively the greatest art heist in history."⁴⁵ She also agreed that this activity is "[p]erpetrated by respectable-seeming corporate entities

³⁷ See House Science, Space and Technology Committee Hearing titled *Trustworthy AI: Managing the Risks of Artificial Intelligence*, Sept. 29, 2022, (a video of that hearing is available at <https://www.youtube.com/watch?v=BcdqyETo4Zg&t=3769s>); Senate Judiciary Subcommittee on Privacy, Technology and the Law Hearing titled *Oversight of A.I.: Rules for Artificial Intelligence*, May 16, 2023 (a video of that hearing is available at <https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence>); House Judiciary IP Subcommittee hearing titled *Artificial Intelligence and Intellectual Property: Part I: Interoperability of AI and Copyright Law*, May 17, 2023, (a video of that hearing is available at https://www.youtube.com/watch?v=Mm1NQ_Kqumw). See also the June 22, 2023 letter from Senators Markey and Peters calling for the U.S. Government Accountability Office to conduct a detailed technology assessment of the potential harms of generative artificial intelligence and how to mitigate them, including an assessment of the extent to which leading generative AI model providers follow practices for "documenting and disclosing training data (including copyrighted and private consumer data)." The letter is available at https://www.markey.senate.gov/imo/media/doc/senator_markey_letter_to_gao_on_generative_ai_-_062223pdf.pdf.

³⁸ See Schumer's SAFE Innovation Framework, available at

https://www.democrats.senate.gov/imo/media/doc/schumer_ai_framework.pdf.

³⁹ See the video of the May 16, 2023 hearing of the Senate Judiciary Subcommittee on Privacy, Technology and the Law titled *Oversight of A.I.: Rules for Artificial Intelligence*, available at <https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence>, starting at 1:07:44.

⁴⁰ See also Van Otterloo, "A checklist for auditing AI systems," ICT Institute, Dec. 18, 2022, available at <https://ictinstitute.nl/a-checklist-for-auditing-ai-systems/>, which notes that intellectual property issues should be considered when auditing data sources.

⁴¹ See <http://garymarcus.com/index.html> for information about Professor Gary Marcus.

⁴² See <https://twitter.com/GaryMarcus/status/1655244591870914565?s=20>.

⁴³ See <https://twitter.com/GaryMarcus/status/1384173525368393736?s=20>.

⁴⁴ See <https://www.linkedin.com/in/timnit-gebru-7b3b407/> and <https://www.dair-institute.org/team> for information about Ms. Gebru.

⁴⁵ See <https://twitter.com/timnitGebru/status/1657891333166743553?s=20>.

backed by Silicon Valley venture capital. It's daylight robbery.”⁴⁶

To address this, the U.S. should promote a policy that requires AI developers and deployers to first obtain appropriate licenses and authorizations in connection with any materials they desire to ingest for purposes of AI development, and before deploying any AI system that was developed using such material. Those policies should also require record-keeping and audits of such authorizations and licenses.

*Vocal and Likeness Cloning are a National Security Risk.*⁴⁷ While AI vocal and likeness cloning have been disruptive for our members,⁴⁸ the use of these AI technologies has broader national security implications. Vocal cloning technology has gone viral, with several Discord servers and social media posts dedicated to posting AI vocal clone models and instructing people on how to create AI vocal clone models of specific people.⁴⁹ On just one such Discord server, over 30 AI voice models of specific people were posted in just one 24-hour period,⁵⁰ and in the past couple of months, just one account on another prominent service has posted over 800 AI voice clone models.⁵¹ Similarly, the number of widely available services offering the ability to make deepfakes is expanding.⁵² While there are hundreds of unauthorized AI voice models of popular singers, rappers, and actors, and of unauthorized image models of various celebrities, one can also easily find publicly available AI voice or image models of political or corporate figures, such as President Biden,⁵³ Donald Trump,⁵⁴ Barack Obama,⁵⁵ Benjamin Netanyahu,⁵⁶

⁴⁶ See <https://twitter.com/timnitGebru/status/1657891342385831936?s=20>.

⁴⁷ Responsive to question 7.

⁴⁸ See, e.g., Harrison, Ellie, “This is going to disrupt the music industry completely: AI-generated Kanye West verse goes viral,” Independent, Mar. 27, 2023, available at <https://www.independent.co.uk/arts-entertainment/music/news/ai-kanye-west-voice-verse-music-b2308511.html>; Coscarelli, Joe, “An A.I. Hit of Fake ‘Drake’ and ‘The Weeknd’ rattles the Music World,” New York Times, Apr. 19, 2023, available at <https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html?searchResultPosition=6>; Coffee, Patrick, “‘Deepfakes’ of Celebrities Have Begun Appearing on Ads, With or Without Their Permission,” Wall Street Journal, Oct. 25, 2022, available at <https://www.wsj.com/articles/deepfakes-of-celebrities-have-begun-appearing-in-ads-with-or-without-their-permission-11666692003>.

⁴⁹ See, e.g., AI Hub Discord server, <https://discord.com/channels/1089076875999072296/1099149801054019604> (over 152,000 members on this server) and AI World Discord server, <https://discord.com/channels/1090114423361916948/1090351949657944216> (over 76,300 members on this server); <https://www.reddit.com/r/VocalSynthesis/> (5.9K members, top 10% ranked by size on Reddit); <https://huggingface.co/QuickWick/Music-AI-Voices/tree/main>.

⁵⁰ See AI Hub Discord server, <https://discord.com/channels/1089076875999072296/1099149801054019604>.

⁵¹ See <https://huggingface.co/QuickWick/Music-AI-Voices/tree/main>.

⁵² See, e.g., Sha, Arjun, “10 Best Deepfake Apps and Websites You Can Try for Fun,” Beebom, Jun. 19, 2023, available at <https://beebom.com/best-deepfake-apps-websites/>.

⁵³ See <https://discord.com/channels/1089076875999072296/1111937280257564782/1111937280257564782>; <https://drive.google.com/file/d/13OGnGVkt1a9yu7zeNnH-a58JfVjspsWq/view>; QuickWick/Music-AI-Voices at main (huggingface.co). Note several of these voice clone models are based on SVC or RVC vocal cloning technology, which has its origins in China. See Zhou, Viola, “Fans in China used AI to deepfake a pop star’s return to music,” restofworld, Jun. 1, 2023, available at <https://restofworld.org/2023/deepfake-pop-songs-stefanie-sun-ai/>.

⁵⁴ See <https://discord.com/channels/1089076875999072296/1112519381314981960/1112519381314981960>; <https://drive.google.com/file/d/1bLL2w6GOO5Pb6dK-IBIRbhKhqM9zUPgt/view>; <https://huggingface.co/QuickWick/Music-AI-Voices/tree/main/Trump%2068k>.

⁵⁵ See <https://blog.deepfakesweb.com/obama-deepfake/>.

⁵⁶ See <https://discord.com/channels/1089076875999072296/1121224280475181157/1121224280475181157>; <https://drive.google.com/file/d/1yPoKXPFw0gZD8IS1AVHl1bk6veWj5kJQ/view>; QuickWick/Music-AI-Voices at main (huggingface.co).

Yoon Suk Yeol,⁵⁷ and Elon Musk.⁵⁸ In addition, several services and apps have recently come online where one can create a vocal or likeness clone of anyone’s voice or likeness without the need for any particular technical knowledge.⁵⁹

These tools can be used by anyone to create voice or likeness models of other political or corporate figures, or everyday Americans, and those models can be used to create and spread disinformation that sounds or looks eerily realistic.⁶⁰ As one commentator noted, this type of AI, along with other AI technology, has become so dangerous that it should be central to our foreign policy.⁶¹

Advancing equity and strengthening civil rights⁶²

Ethical use of AI can help foster creativity and expand economic opportunities in the creative industries, including for individual creators, by, for example, its ethical use for ideation, its authorized use to create versions of a person’s voice in multiple languages to increase the audience for the particular audio using that person’s voice, and its use to reduce the time to edit and finalize human creative expression.

However, some uses of AI, including AI vocal cloning in the music sector, have a disproportionate and negative impact on hip-hop and rap recording artist community, as several unauthorized AI cover songs using the recording artist’s voice (or another’s AI voice overdubbed on the recording artist’s music) have gone viral. While some hip-hop artists “may find ways to . . . remain beneficiaries of their own inventions,” it is also possible, as one commentator noted, that this AI vocal cloning technology “will be a watershed moment when new technology, by allowing the dissemination of forgeries, works to separate artists from their art, thereby making music less democratic, not more.”⁶³ Just as with hip-hop music, misappropriation of an individual’s work, whether in the music, writing, art, or sports, in AI systems that generate output that competes with or dilutes that individual’s work, disenfranchises those individuals by unfairly trading on their lived experiences, their cultural contributions to our society, and their identity, all without authorization or compensation.

Promoting Economic Growth and Good Jobs⁶⁴

As noted previously, the creative community contributes substantially to good, high paying U.S. jobs, and to the U.S. economy. Because of the importance of these jobs and the copyright sector to U.S. economic

⁵⁷ See [https://huggingface.co/QuickWick/Music-AI-Voices/blob/main/Yoon%20Suk%20Yeol%20\(South%20Korea%20President\)%201k%20Epoch/Yoon%20Suk%20Yeol%20\(South%20Korea%20President\)%201k%20Epoch.zip](https://huggingface.co/QuickWick/Music-AI-Voices/blob/main/Yoon%20Suk%20Yeol%20(South%20Korea%20President)%201k%20Epoch/Yoon%20Suk%20Yeol%20(South%20Korea%20President)%201k%20Epoch.zip).

⁵⁸ See <https://huggingface.co/QuickWick/Music-AI-Voices/tree/main/Elon%20Musk%2099K>. See also <https://blog.deepfakesweb.com/elon-musk-deepfake/>.

⁵⁹ See, e.g., <https://beta.elevenlabs.io/>, <https://voice.ai/voice-cloning>, <https://voqul.io/>, <https://musicfy.lol/>, etc.

⁶⁰ Note this vocal clone technology has already been used to create an AI vocal clone of a loved one’s voice to scam individuals. See Verma, Pranshu, “They Thought Loved Ones Were Calling For Help. It Was an AI Scam,” Washington Post, Mar. 5, 2023, available at <https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/>.

⁶¹ See Wright, Robert, “Opinion: AI has become dangerous. So it should be central to foreign policy,” Washington Post, Jun. 7, 2023, available at <https://www.washingtonpost.com/opinions/2023/06/07/ai-foreign-policy-us-china-cold-war/>.

⁶² Responsive to questions 9, 10, 12, and 13.

⁶³ Carson, A.D. et al., “Will AI inspire hip-hop artists – or displace them?” Washington Post, Jun. 2, 2023, available at <https://www.washingtonpost.com/made-by-history/2023/06/02/hip-hop-rap-ai/>.

⁶⁴ Responsive to questions 18, 20, and 22.

growth, care should be taken so that any AI development and deployment expands these jobs, and does not unfairly undermine or threaten them.

As discussed above, ethical development and use of AI can spur economic growth by expanding the range of personal creativity, reducing the time to edit and finalize a work, helping with ideation and fine-tuning of a work, as well as the marketing and promotion their work. However, jobs in each of these areas are under threat by unethical and unfair AI systems development and deployment that ingest copyrighted works without authorization, trade on a person's identity and good will without their consent, and provide output that competes with or dilutes the person's creative expression, good will, or identity.⁶⁵

To mitigate against these harms, U.S. policy should promote strong protection of human rights, including copyright, as a key pillar of its AI policy. This includes requiring AI developers to obtain appropriate licenses or authorization to use any ingestion materials that are copyrighted or that implicate a person's rights of publicity or privacy, requiring adequate record-keeping and auditing, and ensuring that appropriate transparency is maintained. In addition, the U.S. should promote and invest in proper education on the ethical development and deployment of AI, including education on copyright and the importance of seeking authorization before ingesting copyrighted materials or materials that implicate a person's rights of publicity or privacy.

Innovating in public services.⁶⁶

We are concerned with the Federal Government using generative AI in Federal Government services and operations if those generative AI systems were developed by ingesting copyrighted works or works that implicate a person's rights of publicity or privacy without authorization. If the Federal Government uses generative AI, it should engage in extensive due diligence prior to deploying any such AI to ensure appropriate consents and authorizations were obtained in connection with the AI system's design, development, and deployment, as well as to ensure that the system is free from bias and has been designed in a manner to mitigate infringement of a person's fundamental human rights.

Additional Input.⁶⁷

The U.S. should address concerns with data voids or the lack of data quality in AI systems by encouraging and prioritizing the use of quality materials for ingestion into AI models. To encourage the use of quality materials for ingestion, policymakers should encourage voluntary, marketplace licenses among the creators and rights owners, on the one hand, and those AI developers that seek to use such materials for their AI development and deployment purposes, on the other.

Failure to respect the rights of the creators and owners of quality materials, or the rights of performers

⁶⁵ A recent GAO Science and Tech Spotlight on Generative AI notes the economic issues that arise if generative AI systems are trained on copyrighted, proprietary, or sensitive data, without the owner's permission or subject's knowledge. See "Science and Tech Spotlight: Generative AI", Jun. 13, 2013, GAO-23-106782, available at <https://www.gao.gov/products/gao-23-106782>. See also "Generative Artificial Intelligence and the Law," Congressional Research Service, May 11, 2023, available at <https://crsreports.congress.gov/product/pdf/LSB/LSB10922>, which describes the legal concerns with copyright infringement by generative AI.

⁶⁶ Responsive to question 27.

⁶⁷ Responsive to question 29.

and athletes in their name, image, likeness, or voice, will ultimately lead to the degradation of materials available for AI ingestion, loss of human content producers, and ultimately to poor performance of AI models themselves. As Senator Klobuchar pointed out at the May 16, 2023 hearing, “[u]nless you start compensating for everything from books, movies, yes but also news content, we are going to lose any realistic content producers.”⁶⁸ Prominent academics have raised similar concerns with the degradation of AI systems that are partially trained on AI created output.⁶⁹ Simply put, the quality of our nation’s cultural output will decline if we continue to rely on AI systems made by AI developers and deployers who refuse to pay for quality materials. This will harm not only the creative community, but will also lead to lasting negative impacts on our nation’s cultural identity.

3. Conclusion

We thank OSTP for the opportunity to provide these comments and look forward to working with you and other policymakers as our national AI policy is further developed.

⁶⁸ See the video of the hearing, available at <https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence>, starting at 1:16:17. The loss of truly new, quality, human creative expression will accelerate as those individuals, having to compete with AI systems trained on those individuals’ works without any compensation to those individuals, are forced to leave the creative fields in order to make a living.

⁶⁹ Martinez, Gonzalo et al., “Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet,” Jun. 8, 2023, available at <https://arxiv.org/pdf/2306.06130.pdf> (“Our results show that the quality and diversity of the generated images [from AI tools] can degrade over time suggesting that incorporating AI-created data [in AI training materials] can have undesired effects on future versions of generative models.”).

Appendix A – List of Signatories

American Association of Independent Music. The American Association of Independent Music (A2IM) is a 501(c)(6) not-for-profit trade organization headquartered in New York City that exists to support and strengthen the independent recorded music sector and the value of recorded music copyrights. Membership currently includes a broad coalition of hundreds of Independently-owned American music labels. A2IM represents these independently owned small and medium-sized enterprises' (SMEs) interests in the marketplace, in the media, on Capitol Hill, and as part of the global music community. In doing so, it supports a key segment of America's creative class that represents America's diverse musical and cultural heritage. Billboard Magazine identified the independent music label sector as 37.32 percent of the music industry's U.S. recorded music sales market in 2016 based on copyright ownership, making Independent labels collectively the largest music industry sector.

American Federation of Musicians. 80,000 musicians comprise the American Federation of Musicians of the United States and Canada (AFM). We perform in orchestras, backup bands, festivals, clubs, and theaters—both on Broadway and on tour. AFM members also make music for films, TV, commercials, and sound recordings. As the largest union of musicians in the world, we have the power to make the music industry work for musicians.

American Photographic Artists. The American Photographic Artists (APA) is a leading nonprofit organization run by, and for, professional photographers since 1981. Recognized for its broad industry reach, APA works to champion the rights of photographers and image-makers worldwide. APA's mission is to advocate, educate and elevate the professional photographic community.

American Society of Composers, Authors, and Publishers. The American Society of Composers, Authors and Publishers (ASCAP) is a membership association of more than 920,000 songwriters, composers and music publishers, and represents some of the world's most talented music creators. Founded and governed by songwriters and composers, it is the only performing rights organization in the U.S. that operates as a not-for-profit. ASCAP licenses a repertory of over 18 million musical works to hundreds of thousands of businesses that use music, including streaming services, cable television, radio and satellite radio and brick and mortar businesses such as retail stores, hotels, clubs, restaurants and bars. ASCAP collects the licensing fees; identifies, matches and processes trillions of performances every year; and returns nearly 90 cents of every dollar back to its members as royalties. The ASCAP blanket license offers an efficient solution for businesses to legally perform ASCAP music while respecting the right of songwriters and composers to be paid fairly. ASCAP puts music creators first, advocating for their rights and the value of music on Capitol Hill, driving innovation that moves the industry forward, building community and providing the resources and support that creators need to succeed in their careers.

Artist Rights Alliance. Artist Rights Alliance (ARA) is an artist-led advocacy organization fighting for the rights of musicians, songwriters, & performers in the digital economy.

Association of American Literary Agents. The Association of American Literary Agents (AALA) is the preeminent organization of and for professional literary agents in the United States. Since its founding in 1991, the AALA has been a leading force in furthering the interests of agents, authors and other rights holders. Through regular educational programming, community-building initiatives, and advocating for agents and authors alike, the volunteer-run organization is dedicated to helping its members maintain and broaden their professional skills in a fast-changing publishing environment.

Association of Independent Music Publishers. The Association of Independent Music Publishers (AIMP) was formed in 1977 by a group of Los Angeles music publishers, and has local chapters in Los Angeles, New York, Nashville and Atlanta. The organization's primary focus is to educate and inform music publishers about the most current industry trends and practices by providing a forum for the discussion of the issues and problems confronting the music publishing industry.

Black Music Action Coalition. Black Music Action Coalition (BMAC) is an advocacy organization formed to address systemic racism within the music business. Our Coalition advocates on behalf of Black artists, songwriters, producers, managers, agents, executives, lawyers, and other passionate industry professionals.

Broadcast Music, Inc. Celebrating over 80 years of service to songwriters, composers, music publishers and businesses, Broadcast Music, Inc.® (BMI®) is a global leader in music rights management, serving as an advocate for the value of music. BMI represents the public performance rights in over 20.6 million musical works created and owned by more than 1.3 million songwriters, composers, and music publishers. The Company negotiates music license agreements and distributes the fees it generates as royalties to its affiliated writers and publishers when their songs are performed in public. In 1939, BMI created a groundbreaking open-door policy becoming the only performing rights organization to welcome and represent the creators of blues, jazz, country, and American roots music. Today, the musical compositions in BMI's repertoire, from chart toppers to perennial favorites, span all genres of music and are consistently among the most-performed hits of the year.

Christian Music Trade Association. The Christian Music Trade Association (CMTA) is a non-profit organization that exists to build community and cooperation among Christian & Gospel music industry leadership in order to address mutual issues and to maximize Christian/Gospel music's impact on culture. The CMTA supports and promotes all styles of gospel music including pop, black gospel, hip hop, rock, country, southern gospel and more.

Department for Professional Employees, AFL-CIO. The Department for Professional Employees, AFL-CIO (DPE) is a coalition of 24 unions representing over four million professional and technical union members. DPE affiliate unions represent professionals in over 300 occupations in education and healthcare; science, engineering, and technology; legal, business, and management; media, entertainment, and the arts; and public administration.

Folk Alliance International. Folk Alliance International (FAI) is the largest and leading non-profit organization for folk music with a broad coalition of over 3,000 members as well as 50 smaller nonprofits administered via its group exemption, representing the US and extended global community of artists, agents, managers, labels, publicists, arts administrators, venues, festivals, and concert series presenters. Founded in 1989 and based in Kansas City, Missouri, FAI is committed to serving, strengthening, and engaging the folk music community through preservation, presentation, and promotion, and is an NEA-designated National Arts Service Organization, producing the world's largest annual folk music conference as well as the International Folk Music Awards, an Artist in Residence, the Ethno USA gathering, year-round professional development webinars, the Folk DJ Chart, and Global Summits that connect professionals around the folk music world. FAI defines folk broadly as "the music of the people" and supports a diverse array of subgenres including Appalachian, Americana, Blues, Bluegrass, Celtic, Cajun, Global Roots, Hip-Hop, Old-Time, Singer-Songwriter, Spoken Word, Traditional, Zydeco, and various fusions.

Gospel Music Association. Founded in 1964, the Gospel Music Association (GMA) serves a richly diverse community of creatives and professionals within the Christian and Gospel music industry. Through the GMA Foundation and the GMA Gospel Music Hall of Fame, we aim to preserve the legacies of our genre's trailblazers while celebrating the work of today's artists through worldwide events like the GMA Dove Awards.

Graphic Artists Guild. The purpose of the Graphic Artists Guild is to promote and protect the social, economic, and professional interests of its members. We are committed to welcoming, serving, and improving conditions for graphic artists at all skill levels while raising standards for the entire industry. In addition to creative professionals, our members include educators, intellectual property lawyers, artist representatives, and others in related and supporting industries.

Major League Baseball Players Association. The Major League Baseball Players Association (MLBPA) is the labor union that represents all professional baseball players signed to a Major League contract with one of the thirty Clubs in Major League Baseball, totaling roughly 1,200 professionals, along with approximately 5,500 Minor League players also employed by Major League Clubs. In addition to representing our members in the negotiation and enforcement of collective bargaining agreements, the MLBPA also oversees MLB Players, Inc., which exclusively administers the group licensing and commercial rights of all active Players at the Major League and Minor League levels.

Major League Soccer Players Association. The Major League Soccer Players Association (MLSPA) is a democratic organization run by and for players, and serves as the exclusive collective bargaining representative for all current players in Major League Soccer.

Music Artists Coalition. The Music Artists Coalition (MAC) was founded by music creators and industry leaders to advocate on topics that impact music creators. MAC represents artists and songwriters interests without compromise because music creators should be driving the conversation about the issues that shape their lives. MAC believes artists should have the opportunity to decide how to best protect the fate of their music, their other rights and their fans.

Nashville Songwriters Association International. The Nashville Songwriters Association International (NSAI) is the world's largest not-for-profit songwriters trade association. Established in 1967, the membership of more than 5,000 active and professional members spans the United States and foreign countries. NSAI is dedicated to protecting the rights of and serving aspiring and professional songwriters in all genres of music.

National Football League Players Association. The National Football League Players Association (NFLPA) is the union for professional football players in the National Football League. The NFLPA represents all players in matters concerning wages, hours and working conditions and protects their rights as professional football players. Founded in 1956, the NFLPA was established to provide players with formal representation to negotiate compensation and other terms of a collective bargaining agreement. The union also negotiates and monitors retirement and insurance benefits for former players and enhances and defends the image of players and their profession on and off the field.

National Hockey League Players' Association. The National Hockey League Players' Association is a labor organization representing the collective interests of the players in the National Hockey League.

National Music Publishers' Association. The National Music Publishers' Association (NMPA) is the principal trade association representing the U.S. music publishing and songwriting industry. NMPA

represents publishers and songwriters of all catalog and revenue sizes, from large international corporations to small businesses and individuals. Taken together, compositions owned or controlled by NMPA members account for the vast majority of the market for musical composition licensing in the United States. NMPA protects and advances the interests of music publishers and songwriters in matters relating to both the domestic and global protection of music copyrights before the legislative, judicial and executive branches of the U.S. government.

Recording Industry Association of America. The Recording Industry Association of America (RIAA) is the trade organization that supports and promotes the creative and commercial vitality of music labels in the United States, the most vibrant recorded music community in the world. Our membership – which includes several hundred companies, ranging from small-to-medium-sized enterprises to global businesses – creates, manufactures, and/or distributes sound recordings representing the majority of all lawful recorded music consumption in the United States. In support of its mission, the RIAA works to protect the intellectual property and First Amendment rights of artists and music labels; conducts consumer, industry, and technical research; and monitors and reviews state and federal laws, regulations, and policies.

Screen Actors Guild-American Federation of Television and Radio Artists. The Screen Actors Guild-American Federation of Television and Radio Artists (SAG-AFTRA) is the nation’s largest labor union representing working media artists. SAG-AFTRA represents more than 160,000 actors, announcers, broadcasters, journalists, dancers, DJs, news writers, news editors, program hosts, puppeteers, recording artists, singers, stunt performers, voiceover artists and other media professionals. Its members are the faces and voices that entertain and inform America and the world. SAG-AFTRA exists to secure strong protections for media artists. The professionals represented by SAG-AFTRA invest their entire lives in building their professional careers. While most may never be “famous,” their names, voices, images or likenesses – their very persona – have or will attain commercial value. Misuse of these assets can detrimentally impact a public figure’s career and reputation and, consequently, the economic value in their persona.

SESAC Music Group. SESAC Music Group is a global, multi-line music company that provides a wide range of data, technology and services to publishers, songwriters, composers and creators across every corner of the music industry. With its four divisions including Performing Rights, Music Services, Audiovisual Music and Church Music Resources, SESAC Music Group drives efficiency in licensing for music users and is home to a wide array of world-class music companies. The Performing Rights Division is a leader in the domestic exploitation of performing rights, focusing on the highest quality and most valuable musical works from affiliated writers and composers such as Adele, Bob Dylan, Ariana Grande, Burna Boy, Jack Harlow, George Clinton, Glass Animals, REM and Green Day. The Music Services division is a collection of world-class companies providing a wide range of technology, licensing and administration services for creators, publishers, and businesses that use music. The Audiovisual Music division is a leading global provider of original, high-quality music for feature films, scripted and unscripted television shows, local broadcast television news shows, agencies, and brands. The Church Music Resources division provides affordable and legal content, resources, and media to the global church community.

Songwriters of North America. The Songwriters of North America (SONA) is a membership organization founded in 2015 by a group of songwriters, composers and music industry professionals determined to advocate for themselves in an increasingly challenging digital economy. SONA has since evolved into a

trade association and hub for thousands of engaged, working music creators, representing the boots-on-the-ground songwriters and composers who call making music their job.