Before the
UNITED STATES COPYRIGHT OFFICE
Library of Congress

Notice of Inquiry

Artificial Intelligence and Copyright: Notice and Request for Public Comment

Docket No. 2023-6

REPLY COMMENTS OF UNIVERSAL MUSIC GROUP

Jeffrey Harleston
General Counsel and EVP, Business and Legal Affairs
Alasdair McMullan
Head of Litigation and EVP, Business and Legal Affairs
Universal Music Group
2220 Colorado Avenue
Santa Monica, California 90404

David Kokakis
Chief Counsel
Universal Music Publishing Group
2105 Colorado Avenue
Santa Monica, California 90404

Universal Music Group ("UMG") respectfully submits these reply comments in response to the Copyright Office's Notice of Inquiry ("NOI") concerning artificial intelligence and copyright.

As noted in our initial submission, UMG is a leading global music company[1] and home to iconic imprints such as Blue Note, Capitol, Def Jam, Deutsche Grammophon, Interscope, Island, Motown, Republic, UMG Nashville, Universal Music Latin Entertainment, and Virgin. UMG is also home to Universal Music Publishing Group ("UMPG"), one of the industry's premier music publishing operations that publishes and administers some of the most important musical compositions in modern history. Collectively, UMG owns or controls a catalog of sound recordings and musical compositions of immense artistic, cultural, and economic value.

As noted in our initial comments, we believe there are two fundamental principles that form the foundation of sound policy involving generative AI and copyright. First, copyright protection can and should only extend to expressive content created by human beings. This principle is fundamental to the constitutional underpinning of copyright law. While technology can assist human authorship, it should not and cannot replace it. Second, the use of copyrighted content by AI, whether for training or other purposes, requires permission from the copyright owners, via negotiated licenses. UMG and the entirety of the music industry has an extraordinary record of developing bespoke free-market business partnerships with hundreds of digital partners, both large and small. Importantly, these arrangements respect the rights of artists and songwriters, while fostering innovation and unparalleled consumer experiences and access.

[1] https://www.universalmusic.com/

The nearly 10,000 initial responses to the Office's Inquiry clearly demonstrate the import and interest in this imperative issue. Given the breadth of our initial comments and the relatively short reply period, we intend to address only a few of the key arguments raised by respondents in these comments. Our decision not to address a particular issue should not be interpreted as agreement, disagreement, or lack of interest in the opinions espoused by others in their initial comments. However, we note our strong agreement with the points made in the reply comments filed by the Association of American Independent Music (A2IM), the Recording Academy, and the Recording Industry Association of America (RIAA), who expertly addressed several significant topics we do not address here.

Once again, thank you for undertaking this important study and for the opportunity to share our views with the Office.

FAIR USE

In our initial submission, we addressed in detail why the use of copyrighted music for purposes of training generative AI cannot constitute fair use. As we explained, such use runs afoul of the broader purposes and policies of the fair use doctrine and fails to qualify as fair under the four statutory factors that govern the doctrine's analysis. We do not repeat those arguments here. However, our review of submissions by those who hold the contrary view reveals an inaccurate depiction of the realities of AI systems and corresponding misconceptions about how fair use applies to generative AI. We direct our comments to those issues as applied to copyrighted sound recordings and musical compositions.

Common Ground – Generative AI Requires Massive Copying of Copyrighted Works to Create Uncopyrightable Content

We start with two points of relevant common ground concerning both the input and output of generative AI. First, at the *input* and training stages, there is no dispute that generative AI makes unauthorized copies of copyrighted works on a massive scale. *See, e.g.,* Google at 9 ("If training could be accomplished without the creation of copies, there would be no copyright questions here"; "as a technological matter, copies need to be made"); Andreessen Horowitz at 5 ("the only practical way generative AI models can exist is if they can be trained on an almost unimaginably massive amount of content, much of which … will be subject to copyright"); OpenAI at 12 ("… technical realities require that copyrighted works be reproduced"); Samuel, Sprigman, and Sag at 7 ("the process of training Generative AI models is generally preceded by massive amounts of web scraping that results in the creation of locally stored copies of millions or billions of copyrighted works"). Various commentators protest that AI models do not "store" the copied content on which they train. *See, e.g.,* Meta comments at 6; OpenAI comments at 6; Public Knowledge at 10. Whether true or not, storage is irrelevant. A copyright owner has the exclusive right to make copies of their work. Absent a defense like fair use, unauthorized reproduction of that work constitutes infringement regardless of whether the reproduction is retained. *See Capitol Records, LLC v. ReDigi Inc.*, 910 F.3d 649, 658 (2d Cir. 2018) ("even if ReDigi effectively compensated (by offsetting deletions) for the making of unauthorized reproductions in violation of the rights holder's exclusive reproduction right under § 106(1), nonetheless ReDigi's process itself involves the making of unauthorized reproductions that infringe the exclusive reproduction right unless justified under fair use").

Second, there appears to be no dispute that the *output* of generative AI is generally not subject to copyright protection. *See, e.g.,* OpenAI at 15 (because "human authorship [is] a

condition of copyrightability ... OpenAI does not assert any ownership rights over the output of its generative AI products"); Google at 12 ("works that are generated by AI without cognizable human creative intervention are not copyright-eligible because they do not meet the constitutional requirement of authorship"); TechNet at 8 (copyright only protects "the product of human creativity."). While some disagree over the extent to which human control over AI systems can imbue AI-generated content with copyrightable authorship, the vast majority of commentators agree as a general matter that outputs generated by AI are uncopyrightable, because they were created by machines, not humans. As set forth below, these concessions inform the fair use analysis.

The AI Community's Flawed Fair Use Analysis

Against this backdrop, the AI community collectively asserts the same theory for why the ingestion of billions of copyrighted works is fair use: training on copyrighted works allegedly involves only "extracting" uncopyrightable statistical information and facts from the copied works, and then using the resulting statistical correlations to generate new creative works. For example, Meta explains that the "goal of such models is to simply extract non-expressive facts and statistics from training data (e.g., what characteristics typify a cat) and use them to generate new content (e.g., an entirely new picture of a cat)." Meta at 7. This use is purportedly "transformative" because it "is in service of a non-exploitive purpose: to extract information from the works and put that information to use by enabling the creation of new, non-infringing works," *id.* at 12, rather than any exploitive purpose to "capture and reproduce expressive material from the training data itself." *Id.* at 15.

Other leading AI companies make the same basic argument. OpenAI at 11 explains that an AI "model is not interested in the expressive aspects of individual copyrighted works,"

because training merely "teaches the model to analyze the structure and syntax of language and images in general terms, to discern the statistical relationships between words, shapes, colors, textures, and concepts." Google at 14 insists that generative AI is "engineered to create new works, not to copy or facilitate the copying of existing works," and Microsoft at 3 explains that it is surely fair use to allow "reproduction of a copyrighted work when … the facts, ideas and concepts within the work are used to create something novel." Academics press the same argument. *See, e.g.,* Samuelson at 8 ("Processing in-copyright works to extract information about the original [work] does not infringe because it does not replicate protected expression").

This sanitized description of generative AI is riddled with error. It is premised on the false notion that copying copyrighted works at the training stage is a legally justifiable first step in generating new, non-infringing content. To be sure, the major AI players concede that AI models may occasionally "memorize" training data and produce infringing outputs, but they minimize such instances as rare, anomalous, and unwanted technical glitches produced by assiduous efforts to evade effective guardrails. *See, e.g.,* OpenAI at 10 ("memorization of training data is exceptionally rare"); Google at 13 ("The possibility that AI models can occasionally, despite the best efforts of their developers, output content that replicates existing expression is a bug not a feature"); Anthropic at 3 ("we recognize that determined parties can violate our governing agreement and policies and evade our technological measures to create infringing outputs"). That depiction is a myth.

In its recently filed lawsuit against Anthropic, UMG and other major music publishers identified 500 illustrative instances where simply asking for the lyrics to popular, copyrighted songs generated nearly identical infringing copies. The level of "determination" required to produce copyright infringement is no more involved than simply asking for it. But the problem

goes much deeper than mere replication, because AI is also alarmingly adept at creating unauthorized (albeit uncopyrightable) derivative works.

As noted in UMG's original submission, *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 143 S. Ct. 1258 (2023) reigned in the runaway notion that adding "new expression, meaning or message" to a copyrighted work necessarily constitutes a transformative fair use. The Court warned that such an overly expansive view would allow "transformative use" to "swallow the copyrighted owner's exclusive right to prepare derivative works." *Id.* at 1282. *See also id.* at 1275 ("an overbroad concept of transformative use, one that includes any further purpose, or any different character, would narrow the copyright owner's exclusive right to create derivative works"). The AI community falls precisely into this trap when it insists that it consumes copyrighted training data only for the purpose of generating new creative content with new expressive value. As set out in the Anthropic complaint, asking for lyrics from one song often yields outputs that mixes those lyrics with other, unrelated lyrics from other songs or with lyrics of the AI system's own invention. These mashups and distortions not only invade the songwriters' right to prepare derivative works but wrest control of their reputation and artistic aesthetics away from them. Uncontrolled and unpredictable, such combinations and derivations undermine the integrity of the songwriter's intentions.

Indeed, even what purport to be the most tightly controlled AI services readily fulfill requests to create mashups of copyrighted works that would no doubt transcend anything the copyright owners would entertain. For example, while OpenAI protests that infringing output is "exceptionally rare" because it has installed guardrails for squelching infringement, its ChatGPT chatbot obligingly mixes and matches lyrics of unrelated songs upon request:

combine lyrics of american pie with a wonderful world

Certainly, here's a creative blend of the lyrics from "American Pie" by Don McLean and "What a Wonderful World" by Louis Armstrong:

```css
I see trees of green, red roses too
I watched the Chevy drive to the levee, under skies of blue
And I think to myself, "What a wonderful world"
Them good ole boys singing 'bout days gone by
This'll be the day that I die, and I sigh
And I think to myself, "What a wonderful world"

So bye, bye, Miss American Pie
Drove my Chevy to the levee, but the levee was dry
And them good ole boys were drinking whiskey 'n rye
Singin', "This'll be the day that I die"
This'll be the day that I die
```

combine lyrics of drop it like it's hot with somewhere over the rainbow

Creating a combination of lyrics from "Drop It Like It's Hot" by Snoop Dogg and "Somewhere Over the Rainbow" by Judy Garland might result in an interesting contrast. Here's a playful blend:

```vbnet
Somewhere over the rainbow, way up high
Drop it like it's hot, let the beat touch the sky

Bluebirds fly, and the dreams that you dare to dream
Snoop D-O-Double-G, drop it with the steam

Somewhere over the rainbow, skies are blue
Drop it like it's hot, make the crowd go woo

And the dreams that you dare to, why, oh why can't I?
Drop it like it's hot, under the rainbow in the sky
```

Note that these examples, generated in seconds with single searches, boast that they offer "creative" and "playful blends" of copyrighted materials. They are not protected parodies or critiques, but rather random jumbles of copyrighted works that have no business being combined without authorization of the copyright owners.

It is hard to imagine a greater intrusion into the derivative work right than a service that generates unauthorized combinations and arrangements of musical content simply for the asking. Moreover, discovery of these infringing derivative works may prove difficult, particularly if the original titles of the copyrighted works are not used or there is no attribution identifying the original source material. And, while AI developers are at pains to state that their "purpose" is not to create infringing outputs, *see, e.g.,* Meta at 13 ("the purpose of extracting this information is not to reproduce any aspect of the protected expression found in the training data"), their intentions are irrelevant. Copyright infringement is a strict liability offense. *See, e.g., EMI Christian Music Grp., Inc. v. MP3tunes, LLC,* 844 F.3d 79, 89 (2d 2016). And in truth, there can be no serious question that the AI community is acutely aware of how easily these systems produce infringements, given that a simple search reveals all one needs to know, that there is a chorus of objections from the creative community, and that even the AI community's responses to the Notice of Inquiry acknowledge (albeit grudgingly) the problem.[2]

[2] On this point, UMG disagrees with AI developers' efforts to shift liability for infringing outputs solely to users. *See, e.g.,* OpenAI at 14 ("it is the user who determines whether the output implicates the exclusive rights of a copyright owner"); TechNet at 6-7 ("it is the user of a Generative AI tool who supplies the prompt that dictates the resulting text or image …the law should impose liability on the individuals who use AI-powered image generators to achieve the same result"). Those outputs are enabled by the AI systems' unauthorized and knowing reproduction of copyrighted materials in the first instance. Moreover, the AI developers and companies are also subject to the standards of contributory, vicarious and inducement liability, particularly given their awareness of infringement, their material contribution to that infringement, their ability to control it, and their direct financial interest in the very conduct that produces infringement. *See generally* 3 Nimmer on Copyright § 12.04[A][2]-[3] (2023).

But even when the output is not facially similar to the training materials, the massive copying of musical content at the training stage fails to qualify as a fair use for another reason inherent in the very purposes of the fair use doctrine. The AI community's stated justification for this copying is that it is a necessary step to the creation of "new works." *See, e.g.,* Google at 4 ("Generative AI models can use what they have learned to create new content, such as text, images, music, and computer code"); Meta at 9 ("The singular purpose of these models is to enable people to create new creative output to suit their preferences and needs"); TechNet at 1 ("The technology, at its core, promises to enable all people to create new content in any medium or any language, regardless of their skill level or ability"). This capacity to create new works purportedly supports "copyright's [and the fair use doctrine's] overall purpose of promoting progress in the creation and dissemination of new knowledge so that 'each author [can] build[] on the work of others." Samuelson, Sprigman and Sag at 8-9 (citation omitted).

The fallacy in this argument is that generative AI does not stimulate the creation and availability of copyrightable authorship, but rather propagates uncopyrightable imitations of human authorship, which fall outside the ambit of copyright law and thus do not at all vindicate copyright's purposes. As leading AI developers explain, generative AI models copy copyrighted works not to create new authorship, but only to "emulate all facets of human language." See Meta at 2-3. Even where output looks and feels like human expression, it clearly is not:

> To be sure, when, for instance, an LLM is prompted for facts, it can generate articulate responses that may give the impression that it is retrieving information. But, fundamentally, the model is generating responses based on a statistical estimation of what a satisfactory response should look like. Put simply, it produces an average group of words, pixels, or sounds related to a prompt. Some have referred to this as, not an answer, but merely "answer-shaped."

Google at 4. *See also* Samuel, Sprigman, and Sag at 14 (AI uses what it learns "to confect new digital artifacts"). Whatever the benefit or novelty of faithfully mimicking human expression by

a computer program, **it is no part of the purpose of copyright law**. Yes, authors must be able to build upon the works of authors, but there is no copyright benefit to having a computer simply emulate human expression. Such mimicry is not innovative but quintessentially derivative, and if left unchecked, will threaten to crowd out human authorship altogether.

This reality means not only that the use of copyrighted works for training cannot be transformative, but also that such use invades the potential market for copyrighted works embodied in the fourth fair use factor. While we agree that not every use that intrudes into a copyright market is unfair, synthetic substitutes for copyrighted works are the kind of competition copyright law should protect against. Songwriters should not have to suffer the use of their works to enable the creation of "song shaped" compositions or "digital artifacts" that supplant the real thing. Likewise, musical performing artists should not have to compete with machine-created recordings enabled by the copying of their voices and genuine sound recordings.

Consider the pernicious effects of any such exception to liability where digitally generated outputs compete with human authors whose works were used to train the system to generate those outputs. For example, an advertiser needs lyrics for a jingle or a filmmaker needs music for a portion of a film score. The advertiser or filmmaker can engage established lyricists or composers to supply the lyrics or music or can ask generative AI to supply machine-generated substitutes. If the generative AI was trained on the lyricist or composer's copyrighted works, then copyright owners are being unwillingly conscripted into undermining their own legitimate markets and livelihoods. To be clear, UMG does not object to AI-generated lyrics or music as a general matter, but only when they result from training on copyrighted works without authorization. Surely the purposes of copyright law do not embrace this kind of one-for-one

substitution of computer-generated artifacts for human artistry that result from training on the fruits of that artistry in the first instance. As one commentator who otherwise insists that training on copyrighted works is fair use notes, "generative AI holds the potential to displace significant swaths of the creative ecosystem if deployed irresponsibly." *See* Public Knowledge at 17.

Two other themes in the AI community's fair use analysis warrant comment. First, as part of the argument that training only extracts noncopyrightable content, several commentators liken their use of training data to permissible "intermediate copying" or "reverse engineering" en route to the creation of new works. *See, e.g.*, TechNet at 4-5; Google at 11 (citing *Sega Enters. Ltd. V. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992) and *Sony Computer Entertainment, Inc. v. Connectix Corp*., 203 F.3d 596 (9th Cir. 1999), *cert. denied*, 531 U.S. 871 (2000)). As UMG explained in its initial submission, the intermediate copying cases have no application to the use of copyrighted works for training, because (1) generative AI copies and reproduces the essential creative expression of music at the training stage; (2) generative AI produces either infringing outputs or at best synthetic facsimiles of human expression; (3) prohibiting intermediate copying in *Sega* and *Sony* would have perpetuated unfair monopolies not implicated here; and (4) copying computer software is far more susceptible to fair use than copying music, which lies at the core of copyright protection.

Finally, several commentators insist that training generative AI is fair use because AI systems have "substantial non-infringing uses" within the meaning of *Sony Corp. of Am. V. Universal City Studios, Inc*., 464 U.S. 417 (1984). *See, e.g.*, Google at 14 (generative AI is "capable of substantial non-infringing uses, and the law has long been wary of permitting rightsholders to hold up such technologies merely because they could potentially be used for infringing purposes"); OpenAI at 14 (AI models are "capable of substantial non-infringing uses"

and "used primarily for entirely non-infringing purposes"). Some even liken generative AI to content-neutral devices like cameras. *See, e.g.,* TechNet at 3-4; Meta at 7.

But the analogy to *Sony* or cameras simply does not withstand scrutiny. The VCR recorder in *Sony* was simply a device, an empty vessel that contained no content but was capable of being used for both lawful and unlawful purposes. Generative AI is concededly built upon the creation of copies of copyrighted works it selects for training. Unlike a camera or VCR, generative AI is "pre-loaded" by the developer with copyrighted content, and unlike a camera or VCR, AI uses that copyrighted content to generate its own (uncopyrightable) synthetic content. *Sony* has no application to technologies of this nature, which are themselves built upon unauthorized exploitation of copyrighted works and where their proprietors themselves engage in unauthorized copying in the first instance. *See Spanski Enters. v. Telewizja Polska, S.A.*, 883 F.3d 904, 913 (D.C. Cir. 2018) (rejecting video on-demand service's substantial non-infringing use argument where defendant's "own use of [its video on-demand service] to communicate infringing performances amounted to actionable conduct under the Copyright Act").

TRANSPARENCY AND TRADE SECRETS

In resisting the call for transparency, some in the AI community argue that disclosing the identity of training data will jeopardize their valuable confidential information and trade secrets. *See, e.g.*, Google at 12; TechNet at 12. But their description of the sheer breadth of that data and the voracious way in which it is captured is entirely at odds with the very notion of a trade secret.

Trade secrets are generally defined as closely guarded confidential information, developed at great effort and exclusively known to their purported proprietor. For example, the federal Defense of Trade Secrets Act (which is highly similar to most state statutes) defines a trade secret as information whose owner has taken reasonable measure to keep in confidence and

which "derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable through proper means by, another person who can obtain economic value from the disclosure or use of the information." 18 U.S.C. § 1839(3). State common law definitions are even more demanding. New York assesses a claim of trade secrets under the following factors:

> (1) the extent to which the information is known outside of [the] business; (2) the extent to which it is known by employees and others involved in [the] business; (3) the extent of measures taken by [the business] to guard the secrecy of the information; (4) the value of the information to [the business] and [its] competitors; (5) the amount of effort or money expended by [the business] in developing the information; (6) the ease or difficulty with which the information could be properly acquired or duplicated by others.

Ashland Management v. Janien, 82 N.Y.2d 395, 407 (1993). Trade secrets, in other words, must be secret, must be exclusive and must reflect concerted development efforts.

Contrast these restrictive definitions of a trade secret to the apparently unlimited and indiscriminate capture of training data:

- Generative AI is "trained on an almost unimaginably massive amount of content, much of which (because of the ease with which copyright protection can be obtained) will be subject to copyright. For example, large language models are trained on the entire corpus of the written word." (Andreessen Horowitz at 5)

- "Most LLMs are trained on a wide variety of publicly available online data, including web-crawled data, rather than on 'offline' datasets that are prospectively compiled and documented … the source of much of the training, validation, testing, and input data is the massive volume of content available on the entire open World Wide Web." (Google at 11).

- "The breakthroughs in the development of large-scale AI (or foundational) models such as Open AI's Chat GPT require training using vast amounts of data." (Microsoft at 5).

- "Today's most capable LLMs are trained on datasets containing trillions of words and billions of images." (OpenAI at 12).

The AI community speaks with one voice in stressing that there is essentially no limit to the amount of training data it wants. As Microsoft explains, "[d]evelopers of large-scale AI models

therefore optimize for the quantity of data, since the more data available to train a model, the better the performance of the model." Microsoft at 6. In other words, quantity over quality.

Capturing "the entire corpus of the written word," or drawing as much as possible from "the massive volume of content available on the entire open World Wide Web" cannot qualify as a trade secret. That data is publicly available and accessible to any webcrawler and cannot and should not qualify as the exclusive and "secret" property of any one AI developer. Moreover, copying data that contains copyrighted content – or worse, pirated content – is inherently unlawful and infringing. Trade secret protection surely cannot be used to shield infringement from detection. Consider the perspective of the owner of a copyright in musical or other artistic content. What is the point of owning a copyright if a third-party technology company can copy and exploit that content in secret and then claim proprietary "trade secrets" rights in its unauthorized use of those copyrights?

UMG acknowledges that there may be curated and lawful datasets that warrant some measure of protection from public disclosure. But those instances can be managed by procedures that allow copyright owners to identify their works in training data without forcing any broader public disclosure. However, a blanket treatment of training data as "confidential" or trade secrets is both bad policy and legally unjustifiable.

COPYRIGHT ENFORCEMENT AND INNOVATION

A consistent theme throughout many comments submitted by AI companies was the assertion that requiring companies to license copyrighted content for use as AI training material (i.e., to act lawfully) would "inhibit innovation" and "… jeopardize future investment, along with U.S. economic competitiveness and national security." Andreessen Horowitz at 6. One commenter expressed: "There is a very real risk that the overzealous enforcement of

copyright when it comes to AI training—or the ad hoc limitation of the fair use doctrine that properly protects AI training—could cost the United States the battle for global AI dominance." Ibid at 8.

Michael Hiltzik of the *Los Angeles Times* rephrased their unfounded assertion this way: "If the use of copyrighted materials without authorization from the owner is prohibited in these circumstances, investors in generative AI have already "staked so much on an erroneous legal interpretation that we should just give them a pass."[3] Their argument suggests a belief that America is simply incapable of innovating if the law is enforced. A business will certainly have more money to invest if they do not pay for the products, labor, utilities, and other inputs that make their business a business. *But that does not excuse a failure to do so.*

This argument is specious and disproven by the similar confrontations between disruptive technologies and copyright law in the past.

Strong Copyright Protection Fuels and Spurs Innovation. History has demonstrated time and again that robust copyright protection *enables and fuels* innovation. The same reflexive mantra that "enforcing copyrights will impede innovation" was invoked when technology companies argued that copyright law should not restrict peer-to-peer file "sharing." In an Amicus Curiae filed in the landmark *MGM v. Grokster* case, the National Venture Capital Association wrote: "[A ruling against Grokster] would have a devastating impact on the development of legitimate and valuable new products and services for consumers."[4]

---

[3] "Column: AI investors say they'll go broke if they have to pay for copyrighted works. Don't believe it." Los Angeles Times. November 16, 2023.
[4] Brief of the National Venture Capital Association as Amicus Curiae for Respondents at 6, *Grokster*, 535 U.S. 913 (No. 04-480), 2005 WL 497759

But history has clearly proven the opposite: The Supreme Court's unanimous ruling in the *Grokster* case created business certainty and a fair playing field, driving increased investment and strikingly rapid innovation in the digital music space. Since the decision in 2005, advances in music technology – from mp3 downloads to subscription and ad-supported streaming to the licensing of music in the social media space and now in AI – are obvious to even the most casual of observers. Those advancements didn't just "happen"; when the technology and music sectors work together, innovation is spurred. As our CEO, Sir Lucian Grainge, stated recently: "In this dynamic and rapidly evolving market, artists gain most when together we engage with our technology partners to work towards an environment in which responsible AI can take root and grow…This is not a time for passivity. Only with active, constructive, and deep engagement, can we build a mutually successful future together."[5]

The innovation landmarks in the music sector have occurred because intellectual property is a key part of the virtuous cycle that supports innovators, technology users and capital resources.[6] The U.S. has long been a global leader in both innovation[7] and intellectual property protections.[8] Other nations considered the world's most innovative countries practically overlap with those who have the strongest intellectual property regimes.

The WIPO Global Innovation Report notes that the most innovative countries "ensure any national intellectual property (IP) policy is aligned with or even integrated into innovation

---

[5] "Grainge On AI: Engage Actively With Tech." Hits Daily Double. November 16, 2023.
[6] "If We Fail To Protect Intellectual Property, We Destroy Innovation." IAM. February 1, 2018.
[7] Rep. GII 2023 at a Glance The Global Innovation Index 2023 Captures the Innovation Ecosystem Performance of 132 Economies and Tracks the Most Recent Global Innovation Trends. World Intellectual Property Organization (WIPO), 2023.
[8] Rep. 2023 Eleventh Edition International IP Index. U.S. Chamber of Commerce Global Innovation Policy Center (GIPC), 2023.

policy." Countries with strong IP laws can secure stronger trade deals and protect against unfair competition. They witness greater economic development because businesses can be assured protection for their investments. Research and development and the creation of new products are the very heart of innovation, and in nations without strong IP laws, investments in those areas are a risky proposition.

At stake is a very simple principle: if you want to use someone's property to build your business, you need permission to do so. The companies and organizations suggesting that intellectual property guardrails should be "waived" in the interest of AI advancement would be the very first to object if those waivers included patents. They understand, respect, and depend upon vigorous IP enforcement when the IP is their own. That same respect should be shown for other IP rights holders. Doing so will ultimately foster AI innovation and bolster our country's economy.

Generative AI Companies are Capable of Respecting Copyright. Innovative groundbreakers are capable of respecting and licensing copyrighted works. We know this to be true as a number of companies in the space have already licensed our content for other business ventures. Many of the larger corporations involved in generative AI are among the best funded, most technologically brilliant ventures on the planet today. These entities have the means and the technical prowess to create generative AI ventures that respect the property of others. Some of these companies have even stated their willingness to avoid unlicensed training sources.[9]

---

[9] *See, e.g.*, OpenAI at 7 ("We have also engaged in numerous productive dialogues with rightsholders, including authors and music publishers, and asked them to identify sites on the internet that reproduce their copyrighted works. OpenAI has then been able to exclude those sites from being crawled for future training. We have also identified sites that have been identified by rightsholders as hosting infringing content in order to exclude those from being crawled for future training, as well."); "Stability AI Launches Text-to-Music Generator Trained on Licensed Content Via a Partnership with Music Library AudioSparx." Music Business Worldwide. September 14, 2023. ("Unlike some other AI products, the AI

Start-ups (even those with minimal financing) have the capability to do so as well. We license our copyrighted catalogs to a number of fledgling start-ups today and have a proven track record of constructing bespoke arrangements that fit all manner of circumstances.

The Demand for Generative AI Tools that Respect Copyright is High. A legitimate generative AI marketplace is dependent upon safe, transparent, unbiased, and legal content. Companies that train on only licensed materials have a distinct competitive advantage.

In the case of enterprise licensing of AI, many companies are concerned with their own potential liability and actively seek generative AI applications that respect copyright.[10] As the Harvard Business Review warned: "[Businesses] should demand terms of service from generative AI platforms that confirm proper licensure of the training data that feed their AI. They should also demand broad indemnification for potential intellectual property infringement caused by a failure of the AI companies to properly license data input or self-reporting by the AI itself of its outputs to flag for potential infringement."[11] Companies are hungry for legal and ethical generative AI enterprise tools.[12] We have a strong desire to use generative AI tools for various creative and business uses but have struggled to identify "copyright safe" products.

---

algorithm behind Stable Audio was trained on licensed content, via a partnership between Stability AI and music library AudioSparx."); U.S. Congress. Hearing of the Intellectual Property Subcommittee of the Senate Judiciary Committee. "Artificial Intelligence and Intellectual Property – Part II: Copyright." July 12, 2023. Written Testimony of Dana Rao, Executive Vice President, General Counsel and Chief Trust Officer Adobe Inc at 4. ("We chose a path that supports creators and customers by training on a dataset that is designed to be commercially safe. We trained our first Firefly model only on licensed images from our own Adobe Stock photography collection.")

[10] "Does Your Business Need Protection From AI?" Forbes. August 22, 2023.; "15 AI risks businesses must confront and how to address them." TechTarget. June 1, 2023; "The Top Five Real Risks Of AI to Your Business." Forbes. June 22, 2023.

[11] "Generative AI Has an Intellectual Property Problem." Harvard Business Review. April 7, 2023.

[12] U.S. Congress. Hearing of the Intellectual Property Subcommittee of the Senate Judiciary Committee. "Artificial Intelligence and Intellectual Property – Part II: Copyright." July 12, 2023. Written Testimony of Dana Rao, Executive Vice President, General Counsel and Chief Trust Officer Adobe Inc at 5. ("Having a dataset designed to be commercially safe makes Adobe Firefly more attractive to both the creative community – who are eager to use the power of generative AI, but in a way that respects their

Consumers and individual AI users are also wary of opaque AI algorithms that could create infringing works without their knowledge or intent. Major generative AI developers have announced copyright "indemnification" programs for their users,[13] a striking admission of their potential for abuse. Such programs wouldn't be necessary if their users weren't concerned about their own legal liability. Creating a system where users may violate the law could be easily avoided if those companies licensed training materials in the first place.

Respecting Copyrighted Content Leads to a Better Generative AI Product. The adage "garbage in, garbage out" is exceedingly true in the world of generative AI: "The need for *quality* data in generative AI can't be overstated. It's the foundation upon which reliable, effective AI is built" (emphasis added).[14] AI technology will advance when developers work with copyright owners, who have much to offer: access to the highest quality content, rich metadata, metrics, regular safe and secure content deliveries that could protect the model and expedite effective training, exclusive content, beneficial product development expertise, support from artists and songwriters – and much more.

***

Again, we commend and thank you for undertaking this important study. We look forward to working with the Office in the months ahead.

---

fellow creators – and to enterprises who want to use the model to generate brand campaigns and marketing materials without fear of infringing on someone's IP.")

[13] *See, e.g.*, "Copyright Shield." Web log. New Models and Developer Products Announced at DevDay (blog). OpenAI, November 10, 2023.; "Shared Fate: Protecting Customers with Generative AI Indemnification." Web log. AI & Machine Learning (blog). Google Cloud, October 12, 2023.; "Microsoft Announces New Copilot Copyright Commitment for Customers." Web log. Microsoft on the Issues (blog). Microsoft, September 7, 2023.; "IBM Contractual Protections for AI Models." Web log. IBM Announces Availability of Watsonx Granite Model Series, Client Protections for IBM Watsonx Models (blog). IBM, September 28, 2023.