Questions from Senator Tillis
for Jeffrey Harleston
Witness for the Senate Committee on the Judiciary Subcommittee on
Intellectual Property Hearing "Artificial Intelligence and Intellectual
Property - Part II: Copyright"

1. Given generative AI is developing all over the world and countries are responding to it in different ways, are there policies or regulations being adopted elsewhere that you recommend that the U.S. consider or avoid?

    The development of public policy surrounding AI is in its infancy, presenting the U.S. with an important opportunity to lead the world. As the Senate considers 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) and the Ministerial Declaration of the Digital and Tech Ministers' Meeting, (see paragraphs 42, 45 and 47).

    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, the Japanese polices regarding Text and Data Mining exceptions to copyright law as enacted in 2009 and amended in 2018. Another example includes the policies in Singapore, which were enacted in 2021. We would note that the United Kingdom explicitly rejected such policies last year in recognition of the irrevocable harm it would inflict upon their creative industries.

2. A recent survey on how consumers view AI found that most consumers – nearly 80% – believe the use of AI should be explicitly disclosed. Do you agree? Why or why not?

    Yes. Content generated by an AI should be identified as such. Consumers have a right to know that what they are consuming is real. There is exceptionally high risk that content created by generative AI could mislead or deceive consumers, and that risk necessitates disclosure.

    However, merely *disclosing* purely generative AI does not go far enough. If an artist's name, image, likeness, or voice is used without permission, or if an AI engine was trained on unauthorized copyrighted content, mere disclosure does not solve the underlying problem or mitigate the harm. To ensure consumers and fans aren't duped, and artists' livelihood and rights are protected, federal right of publicity legislation and enforcement of copyright law should also be enacted.

    In short, the best way to mitigate the potential risks posed by generative AI is to address and regulate how AI models are trained in the first instance.

3. What are the benefits and disadvantages of requiring an AI company to keep records of everything that is ingested and to make those records publicly available?

    Detailed recordkeeping of data, materials, and information used to train a generative AI engine is essential for the following reasons:

    (1) **Accurate, truthful, and detailed recordkeeping would allow copyright holders to know when their content was used without permission.**

         This is a fundamental tenet of 233 years of American copyright law and jurisprudence. AI developers need permission from copyright owners before their works can be used to train an AI algorithm.

         Recordkeeping is the lynchpin of a copyright owner's ability to enforce their rights. Without the ability to see how an AI was trained, an untenable "Catch-22" exists: copyright law requires rightsholders to prove a registered copyright was violated before they can enforce their rights. Since most AI training sets are not transparent and available to

rightsholders, they must initiate expensive, lengthy litigation in order to reach discovery and prove that their content exists in the training sets. However, without *proof* that a copyright has been violated, rightsholders cannot even initiate the litigation it would take to reach discovery. In other words: rightsholders must sue in order to prove their registered works were used to train an AI, but they *can't* sue unless they can prove their registered works were used. The inability of rightsholders to protect themselves is an unreasonable consequence of the AI era, and Congress should remedy the problem.

(2)   **It would require those who create AI engines to be more thoughtful and careful about the content they use to train their AI engines**. Clarity on recordkeeping requirements would ideally inspire those creating AI engines to seek authorization in the first instance – reducing intellectual property theft and establishing a robust marketplace free of litigation and uncertainty.

(3)   Importantly, **transparency of training information would allow consumers and users of an AI to know what went into the AI that they are using**. Confidence in accuracy, veracity, and trustworthiness of training sets is essential to establish the foundation of a legitimate generative AI marketplace, maximizing potential benefits while mitigating the potential harms.

a. Under what circumstances, if any, should an AI company NOT be required to make its records of everything that is ingested by the AI publicly available?

There is no justifiable circumstance where an AI creator should not have to make their training sets available. While it has been argued that making training set information transparent would threaten "trade secrets" or proprietary information, those arguments do not hold water. The potential for bias, misinformation, and very real harm to American intellectual property owners is too great to justify that position. AI developers have a responsibility to disclose the material used to train their systems to the societies in which they operate.

b. Under what circumstances, if any, should an AI company be required to make its records of everything that is ingested by the AI publicly available?

An AI developer should make everything that is ingested for training of their model available, especially to rightsholders.

4. Do you think that generative AI prompts provided by users are copyrightable? And if so, under what circumstances could they be copyrightable?

A prompt provided by an AI user is, essentially, a text-based work. Therefore, it should be copyrightable to the same extent any text-based work is copyrightable. As the Supreme Court has held:

> "The *sine qua non* of copyright is originality. To qualify for copyright protection, a work must be original to the author. Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity." (*Feist Publications, Inc. v. Rural Telephone Service Company, Inc.,* 1991)

a. Do you think that whether the prompt used is copyrightable or not should impact the copyrightability of the resulting AI output generated as a result of the provided prompt?

No. As a general matter, the copyrightability of a prompt and the copyrightability of an AI output are two separate and distinct inquiries, but which apply the same standards for copyrightability. If an output is generated purely by AI, it fails the copyrightability test as there is no human author.

That being said, many prompts (e.g., "create a picture of a cat on a surfboard") will not contain sufficient originality or creativity needed to be copyrightable. And if the output is generated purely by AI, it fails the copyrightability test as there is no human author.

5. What does the impact of generative AI have on the creative industry? Specifically, what are your thoughts regarding the concern that the proliferation of generation AI will take over jobs?

When copyright is violated, it has a very real financial impact on an artist's livelihood, with ripple effects felt throughout the entire music ecosystem.

Speaking as a music publisher and record label, the existence of content that violates our rights harms our ability to license a creator's music. That means lost U.S. GDP, tax revenue, and an inability to support as many employees, artists, and small businesses. It results in an inability to invest in and support as many new artists, leading to fewer opportunities for songwriters, fewer jobs for producers, studios, instrument and equipment manufacturers, musicians, managers, lawyers, stylists, directors, graphic designers, choreographers, video production crews, caterers, and so on. It means fewer tours supporting fewer venues across America.

In 2021, during the height of the COVID-19 pandemic, the core copyright industries added $1.8 trillion dollars of value to U.S. GDP (accounting for 7.76% of the entire U.S. economy) and employed 9.6 million American workers (accounting for 4.88% of the nation's workforce). The copyright industries also provided the sustenance that fed our societal mental health during that unprecedented and difficult time of isolation, anxiety, and loss.

And it's particularly egregious because it effectively strips the artist/songwriter of their rights of integrity and control: they don't get to choose whether or not their work is used to create new works by others. And since those new works could be political, defamatory, violent, racist, or offensive it's not only troubling, but it can also be tremendously damaging to the artists' brand and commercial potential.

Imagine if someone stole everything of value from a business and used it to start a business to compete against them. That's exactly what's happening with a lot of the larger machine-learning AI models in existence today. This is textbook unfair competition: hundreds of digital services worldwide (as well as individual creators, advertising agencies, small businesses, content creation companies, and others) have followed the law and licensed the work; AI developers should operate under the same rules.

Unless we protect creators in this new era, the long-term consequences on human creativity and our culture could be dire. Some 120,000 songs are uploaded to streaming services *every day*. AI-generated songs will only get easier to make and that number will only rise. Will that make it harder or easier to discover good music created by humans? Will that improve the experience

for music fans? Will this flood of content be better or worse for artists attempting to make a living from their music? We think the answers are clear.

It will take some time for a legitimate, legal marketplace to develop, and you have the power to determine the speed at which it materializes. There is a hunger for marketplace actors who respect copyright. Pending litigation results in marketplace uncertainty stymies forward progress that benefits everyone. There is an urgent need for appropriate "rules of the road" for generative AI and we encourage you to act decisively and without delay.

6.   If a generative AI system is found to infringe a copyrighted work, who should be liable for the infringement – the AI company, the user providing the prompts to the AI tool, or both?

As always, this would be a fact-based analysis, but generally speaking:

TRAINING: If the AI system has been found to be infringing because it trained on copyrighted works without permission, the actors who engaged in that training violation would be liable for infringement. That could be a company or organization, a group of individuals, or a single individual.

PROMPTING: A "prompter" would not necessarily be liable if the content generated in response to their prompt included copyrighted material. A copyright violation occurs if the prompter makes use of the infringing work after its creation. If a prompter *sought* to produce a work through an AI that was infringing, that might render them liable under either direct or contributory infringement principles.

INFRINGING OUTPUTS: In terms of an output, the infringer(s) would be the person or company that is violating the rights guaranteed under U.S. Code Title 17 Sec. 106. If an AI produces something that infringes an underlying work, the question becomes who is copying, performing, transmitting, displaying (etc.) that infringing work.

7.   In your opinion – currently or in the foreseeable future – can AI generated material ever replace the quality of human created work?

No – nothing can replace human creativity, experience, and expression. But in many cases, AI-generated imagery is already indistinguishable from human-created art – especially by the layman – and other forms of media are not far behind. People will always relate best to other people and generative AI will

never be able to make art informed by the human-lived, and often cultural, experience, but the output quality of AI tools will likely equal that of human-created works in the near future, which is why it is critical that Congress take steps to protect human creativity.

Because generative AI is capable of producing remarkably high-quality outputs, it can mislead or deceive consumers into believing something purely generated is real, factual, or has occurred. In the context of music, increasingly high-quality deep fakes or voice clones lead to cultural appropriation, consumer confusion, unfair competition against the actual artist, market dilution, and potentially irreparable damage to the artist's reputation, brand, and livelihood.

8. A balance needs to be struck in terms of how to encourage innovation, how to be responsible, and how to ensure that there is clarity for all using this technology. How do you propose we do this in the copyright space in a way that allows the U.S. to stay competitive and remain the global leader?

    U.S. copyright law has enabled countless technological innovations since it was first established in 1790. Copyright law protects creators' innovations and rather than stymie technological innovation, it has *encouraged and fueled* it. Ethical technological innovation does not tread on the rights of other creators and innovators. Quality content drives a hunger for devices and technology that make the content shine, easier to access, and amazing to experience; and technology companies need quality content to spur a market for their services, software, and devices – just as content owners need technology to help their content reach fans.

    When the law is clear and guardrails against appropriation and abuse are firmly established, technology and content *both* thrive. The proper legal guidelines and guardrails that protect our artists' innovation and expression are essential, and to trample the rights of creators in order to create unprecedented shortcuts for AI developers is unjust.

    The ideal "win-win" scenario is one in which AI companies build their systems using content licensed from creators and content owners in an open market, and all participants benefit from the associated commerce.

9. In the copyright context, what differentiates the technology of generative AI from other machine-aided creativity, such as photography, video cameras,

electronic music, and the like, all of which allow the public to develop and advance knowledge?

Creators use those machines as tools and instruments to produce original works that spring from their original thoughts and imagination. Generative AI is unique in that the creative elements are derivative by nature and born of technology that has been trained on a body of human-created works that have come before. In the case of generative AI, the creative choices are made by the technology, not the human. When using a camera, the photographer chooses what to shoot, how to frame it, where to focus and add light, how to edit it, etc. When a "photograph" is created by an AI, those choices are largely made by machine learning and guided by the materials and information it used to train the algorithm.

Allowing "the public to develop and advance knowledge" inherently suggests human knowledge. While AI can do many amazing things to advance scientific achievement, create efficiencies, and problem solve, we should be careful as a society not to equate the knowledge and learning generated by a machine learning AI with human knowledge and achievement. As computers assume the tasks so many of us struggled ourselves to learn and absorb, our global society should take care to ensure humanity maintains its own genuine intelligence, creativity, and knowledge.

10. What steps can and should the creative community take today to ensure that their work is more easily attributed to them, regardless of whether their work is used for training an AI model? For example, indicating authorship and contact information via the metadata of the author's digital content.

Content created by the music industry is already incredibly rich in detailed metadata and digital information. The global digital music marketplace is reliant on that information for far more than ownership, licensing, and sales information. Metadata and the standardized digital messages sent between business partners may be invisible to the consumer, but together they form the "nervous system" of the vibrant digital music marketplace.

While each company surrounds its content in proprietary metadata, Digital Data Exchange, LLC (DDEX) enables the world's music business partners to communicate that metadata seamlessly. DDEX is a not-for-profit, membership organization comprised by a consortium of media companies, music licensing

organizations, rights owners, digital service providers, and technical intermediaries.

Media-specific industry standards, such as DDEX for music, should continue to be used to transport data about creative works. Information regarding AI is already being implemented into the DDEX system and can help facilitate a legitimate generative AI marketplace for music. Creators and distributors also use industry standard identifiers for contributors and media (such as ISNI, ISRC, ISWC, and UPC for music). These same standards can be used to transmit metadata and content obtained from record labels for AI training purposes. New and developing labeling standards such as C2PA may be used to indicate the provenance of partially AI-created output and identify contributing works in the future, but the space is still developing.

Information regarding AI is already being implemented into the DDEX system and can help facilitate a legitimate generative AI marketplace for music.

11. Are existing laws and regulations sufficient to deal with the issues relating to transparency and record keeping by AI companies?

No. There is a need to future-proof the law to fit a technology evolving and advancing at breakneck speed. There's also a clear problem today where rightsholders have great difficulty enforcing their rights because copyright law is not perfectly fit to allow rightsholders to protect themselves. Moreover, in order to cultivate a lawful, legitimate, trustworthy, and ethical AI marketplace, Congress needs to establish the fundamental "rules of the road" that enable that development.

It will take more than ensuring transparency and detailed recordkeeping to accelerate ethical AI development. Specifically, the U.S. should pursue the following changes to law:

(1) Enact a federal Right of Publicity statute to ensure intellectual property protection for a name, image, likeness, or voice.
(2) Ensure the transparency of AI training materials and allow a rightsholder to view detailed records of training inputs, without having to initiate litigation. Solving the rightsholder litigation "Catch-22" is paramount.
(3) Require labeling of works substantially generated by AI.

(4) Future-proof copyright law to ensure it holds up to an ever-evolving AI technology.

12. Have you reviewed the U.S. Copyright Office's Registration Guidance for "Works Containing Material Generated by Artificial Intelligence" and, if so, what are your views on the guidance?

a. Do you think that the Copyright Office got it right? Are there aspects of the guidance that could stand to be clarified or revised?

We think the Copyright Office established the right principle: that human creativity is necessary for copyright protection and that material that is not generated through the human creative process should not be subject to copyright protection. When registering our works, we believe this means that AI-generated material should be disclosed to the Office when it takes the place of human creative expression.

There is a breadth of AI tools used in the creation of music, and art has always pushed the boundaries of technology. We assume the Copyright Office guidance will develop and evolve over time as the technology advances and the AI market more fully develops.

13. Both the U.S. Patent and Trademark Office and the U.S. Copyright Office have engaged in extensive outreach regarding AI. Have you participated in this outreach and, if so, how did you find it? What more can and should these offices do?

Yes. The music community was represented in both proceedings through our trade associations. In addition, UMG participated in the recent Copyright Office roundtable on Music and AI this past May.

These proceedings have been extremely helpful. While their efforts are ongoing, it is our hope they will serve as needed, necessary voices within Congress and the Administration – creators must be represented when decisions regarding AI policy are made, and we're deeply hopeful they'll actively advocate on our behalf. We're also hopeful they act decisively and quickly, as the speed of AI advancement necessitates it.

14. What are artists saying to you about their voices and music being used without their consent? Is there something that we need to do to protect a person's voice so that others can't use generative AI to exploit their unique qualities?

   Artists are understandably upset that their name, image, likeness, or voice is stolen and used to suggest that they're said, sang, or done something they never did. It's a horrifying violation of privacy, and since it could damage the artist's reputation, it could irreparably damage their career. An artist's voice and persona *is* their livelihood and to steal it – no matter the means – is wrong.

   Adding to the harm, deep-fake and/or unauthorized recordings or visuals of artists generated by AI not only lead to consumer confusion, but they also unfairly compete against the actual artists, diluting the markets for their music and devaluing their brands.

   AI-generated, mimicked vocals trained on vocal recordings extracted from our copyrighted recordings go beyond Right of Publicity violations and concerns about consumer deception, unfair trade practices, and privacy – copyright law has clearly been violated.

   The Senate should enact a federal Right of Publicity statute, ensuring an artist's persona is protected as intellectual property.

15. Should someone be able to use an artist's voice to train an AI model so that an AI system can sing other people's songs or say something new? How is such different from a person imitating another person's sound?

   Yes, but one should only use an artist's voice to train AI with the proper consent.

   An AI voice clone is very different from a human imitation of another human's sound. In terms of economic impact, a generative AI model has no limits and can create an infinite number of recordings that sound exactly like an artist whose voice was stolen. This avalanche of recordings competes against the artist in the marketplace, and every song streamed is money stolen from the artist.

16. Do you have any thoughts on some of the positive ways that AI can be used to support artists?

When used in the service of artists and artistry, generative AI can be powerful and help artists reach new fans in exciting ways. In my testimony, I described how an artist we distribute was able to use a generative AI engine to train on his voice and the intonation of native speakers in six different languages. This allowed him to release a song in seven different languages at the same time, all in his own voice. It's thrilling to think of what such a tool could mean for music and artistry.

Aside from the potential artistic impact, generative AI is also poised to make the life of an artist easier. Touring with greater efficiency, finding new fans, streamlining licensing, monetizing music in new ways never imagined – generative AI could do great things for artists.

17. In your testimony, you outline a number of different ways that you believe AI-generated music violates copyright law. You also point out how often UMG and artists collaborate with other artists and partners to bring music to fans. Whether, it's the recording of a cover of a popular song, sampling of a song's lyrics or composition, or a collaboration on a remix, artists and the industry regularly produce new or reimagined music while following the law.

Could you please explain the legal and appropriate process of creating music involving samples, for instance, and compare that process to what is happening with generative AI music that we've seen in such volume over the last few months?

When an artist wishes to use a sample of a pre-existing recording in a new recording, clearance is generally sought from both the owner of the original (sampled) sound recording and the original (sampled/interpolated) musical work. That clearance is typically handled by the artist's record label and involves obtaining licenses from those original copyright owners in exchange for consideration. There is a well-developed market in clearing and licensing these samples, but these uses are subject to exclusive copyrights. Accordingly, the original creators/owners are entitled to deny these requests.

To date, in the generative AI market there has been simply a wholesale taking of content – entire catalogs – without *anyone seeking permission*. This is a process that cannot continue and is likely to lead to extensive litigation unless Congress asserts the basic tenets of copyright law, making clear that existing law applies to AI developers, just as it applies to everyone else.

18. AI may be increasingly used instead of non-featured vocalists and studio musicians in the making of sound recordings. Assuming this trend continues, what impact will it have on the music industry?

I am aware of a few instances where generative AI has been used instead of background vocalists or studio musicians. It wouldn't be in our interest to use generative AI in that way as it would result in an inability to obtain a copyright for those elements of the completed recording.