DIGIMARC

October 30, 2023

Submitter: Joel Meyer

Organization: Digimarc Corporation

U.S. Copyright Office, Library of Congress

Submitted via regulations.gov

> *Digimarc Corporation's Comments in response to the Copyright Office's Notice of Inquiry and Request for Comments on Artificial Intelligence and Copyright, dated August 30, 2023 (88 Fed. Reg. 59943)*

Digimarc is pleased to provide its Comments in response to the Copyright Office's Notice of Inquiry and Request for Comments on Artificial Intelligence and Copyright, dated August 30, 2023 (88 Fed. Reg. 59943) ("NOI"). Our Comments are presented from the vantage point of the pioneer of a signal processing innovation known as digital watermarking and our almost 30 years of experience deploying digital watermarking for copyright protection, authentication, and anti-counterfeiting.

We first deployed watermarking for copyright communication for images as a plug-in for Adobe Photoshop in 1996. Since then, our innovations and inventions in digital watermarking have been utilized at large scale to deter counterfeiting of currency, to protect motion pictures, to measure television and radio audiences, and to secure the credentials of the majority of U.S. drivers' licenses and other secure credentials in countries around the world.

Our long history in developing and deploying watermarking solutions at large scale enables us to provide unique insight into the questions relating to online copyright management systems and labeling of AI model output. Therefore, we focus our responses on questions 9, 9.1-9.4, 15, 16, 26, 28, 28.1-3, and 29.

Before addressing these questions and to provide context for our comments, we begin with definitions of digital watermarking and a digital watermark.

Definition of Digital Watermark

At a high level, digital watermarking is the science of hiding information about an item in the item itself. The act of hiding that information is known as embedding; the act of discovering it is known as detection.

A digital watermark is:

1. A machine-readable and covert (as opposed to human-readable and human-visible) identifier that, upon embedding, becomes part of the item itself.
2. Because it is covert, it can be ubiquitous, meaning it covers the totality of the item.
3. Because of this ubiquity, as well as the signal itself, it is redundant, meaning that the watermark can be recovered from a portion of an item (e.g., a portion that remains if the content is cropped or clipped).

8500 SW Creekside Place
Beaverton, OR 97008. USA
+1 503 469 4800 +1 800 344 4627

digimarc.com

A digital watermark can be applied to anything that:

1. is digital,
2. is digitally processed, or
3. is made from something digitally processed.

The term "watermark" is sometimes used to refer to a visible icon applied to images, like those visible and readable marks found in the images of stock photography vendors. But such visible icons simply represent a mark applied to a digital item, not a digital watermark. These marks are not covert, and they are not ubiquitous or redundant. As such, they are easily removed and do not facilitate fast, reliable machine detection.

Some companies have referred to the metadata attached to a digital item as a digital watermark. This metadata can be included, for example, in a file header of the file containing the digital item. While the metadata attached to a digital item is machine-readable, the similarity with a digital watermark ends there. Think of the metadata as a digital envelope for the digital item, and this envelope has a wealth of information written on it. It is not ubiquitous or redundant, there is only one envelope, and it – or the information written on it – cannot survive any damage. Moreover, that metadata can (and often is) severed from the digital item, for example when it is posted to a social network or ingested by an editing application. The envelope is ripped open, thrown away, and the item is placed in a new envelope, with brand new instructions. Thus, the metadata does not persist reliably with the content. Further, anyone, at any time, can quite easily change the information found on the envelope.

Digital watermarks may embed metadata in the content, but more typically, they embed flags and one or more identifiers that reference metadata stored separately.

Digimarc Comments

We now turn to questions particularly germane to digital watermarking.

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

Whether an opt-in or opt-out system is adopted, there is a critical need for an efficient and automated means of communicating copyright information to assess whether the copyright owner has authorized use of a work for training of AI models and under what, if any, conditions. As question 9.1 suggests, there are a variety of contexts in which a work may be used to train an AI model. Uses may be authorized or not, depending on several factors. Digital watermarks communicate copyright management information (CMI) within the content itself as well as provide references to descriptive metadata within a registry, including rights holder information and usage permissions. Since they convey CMI within the content, watermarks provide a more reliable way to provide opt outs for AI training than other ways of conveying rights holder authorization, such as attaching metadata to files or placing machine-readable text on web sites.

> *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?*

A digital watermark can provide one or more technical flags indicating permitted or unauthorized uses as well as a reference to metadata stored in a registry. The conveyance of the opt out preference in a digital watermark, embedded within the content, enables that preference to be tailored to the content and reliably persist with it when distributed. The digital watermark, thus, remains embedded within the content through and up to the point of ingestion for AI model training. In addition to conveying technical flags, a digital watermark also provides a reference to metadata, including CMI stored in a registry. This metadata remains linked to the content and can provide additional detail about the work, its creator and owner, and information enabling an individual or organization to obtain rights.

Other means of conveying opt out preferences include attaching metadata to the file and placing machine-readable text on web sites. The metadata attached to a file provides a complementary way to convey CMI. However, this metadata is often stripped in routine processing of content, such as posting to a social network or ingesting in an image-editing application. Metadata that is not persistently linked to the content via a digital watermark is easy to modify or swap with other information. In contrast, digital watermarks are embedded within the content of the audiovisual work itself in a manner that is covert and ubiquitous. These properties enable automated and reliable detection of CMI through the use of digital watermark detection software.

An industry standards body, C2PA[1], has included an opt out mechanism for AI model training in its standard for associating provenance and authenticity information with content. The C2PA specification describes how to attach metadata, called a manifest, to files to detect tampering of content. The manifest may also include an opt out indicating whether a media file or stream may be used as part of data mining or an AI or Machine Learning (ML) workflow.[2] This opt out is referred to as a "Do Not Train" credential.[3] Digimarc is working with C2PA on the use of digital watermarks to bind the manifest to content to address the problem of modification, swapping and removal of the manifest.

The provision of an opt out in machine readable text on a website is useful but insufficient to convey opt out preferences to AI model providers and suppliers of training data. This approach is prescribed in the European Union's Directive on Copyright in the Digital Single Market as a means for rightsholders to opt

[1] The Coalition for Content Provenance and Authenticity (C2PA) addresses the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content. See https://c2pa.org/

[2] See section 19.21. Training and Data Mining of the C2PA Specification https://c2pa.org/specifications/specifications/1.3/index.html

[3] See Content Authenticity Initiative blog post April 3, 2023 https://contentauthenticity.org/blog/meeting-the-moment-with-c2pa-and-firefly

out of the use of their works in text and data mining.[4] This approach has also been suggested as a way to enable websites to opt out of being scraped by a web crawler seeking data for use in training AI models.[5]

We agree that machine-readable means are necessary to convey opt out preferences. However, providing such machine readable-text on a website does not protect audiovisual works after they have been distributed in other channels or removed from the website by third parties. Moreover, it makes it difficult for the rightsholder to identify CMI for particular audiovisual works. Rightsholders publish and distribute their works via their website, and also via channels that include but are not limited to social media channels, PR agencies, vendors, international media organizations, event organizers, and consultants. Given the realities of content publication and distribution, digital watermarks are unique in their ability to address the challenge of conveying rightsholder preferences reliably and persistently with content and with the specificity required to tailor the preferences to individual works.

> *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?*

The volume of works used in training, coupled with the challenge of determining permissions, necessitate the use of automated and reliable means to convey CMI. Digital watermark detection applied at acquisition of content (e.g., from internet sites) and at input to training enables efficient detection of permissions and facilitates getting consent from the rights holder.

Digital watermarks provide a particularly effective way of conveying CMI in the context of AI model training because they enable the generative AI provider to detect CMI automatically, efficiently, and reliably. As such, they enable those training AI models to reliably and feasibly obtain ex ante consent from copyright owners. To facilitate detection, generative AI systems can easily integrate watermark detection at their inputs. Watermark detection is efficient, as it can be implemented as a filter to screen files as they are acquired or ingested. Watermarks are reliable because they apply across file formats and format conversions, remain persistent within content whether or not metadata has been stripped from a file, and survive signal processing operations that are content preserving. Further, watermark detection is reliable as it outputs a deterministic data payload, at an extremely low false positive rate. This is in contrast to other means of associating rights holder information, such as fingerprinting, which is based on a probabilistic (and, therefore, inherently imperfect) match of features of a content item.

---

[4] The European Union's Directive on Copyright in the Digital Single Market provides for two copyright exceptions or limitations for text and data mining (which may be used in the training of generative AI systems): one for purposes of scientific research and one for any other purpose. The latter is available only to the extent that rightsholders have not expressly reserved their rights to the use of their works in text and data mining. The Directive explains that: "In the case of content that has been made publicly available online, it should only be considered appropriate to reserve those rights by the use of machine-readable means, including metadata and terms and conditions of a website or a service." Article 4 of the Directive states: "The exception or limitation … shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online." See Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, 2019 O.J. (L 130), https://eur-lex.europa.eu/eli/dir/2019/790/oj.

[5] See Emilia David, Now you can block OpenAI's web crawler, The Verge (Aug. 7, 2023), https://www.theverge.com/2023/8/7/23823046/openai-data-scrape-block-ai; Melissa Heikkilä, Artists can now opt out of the next version of Stable Diffusion, MIT Tech. Review (Dec. 16, 2022), https://www.technologyreview.com/2022/12/16/1065247/artists-can-now-opt-out-of-the-next-version-of-stable-diffusion/ See also, Garling, How to know if your images trained an AI model (and how to opt out) https://www.makeuseof.com/how-to-know-images-trained-ai-art-generator/ MakeUseOf.com (Jan. 27, 2023).

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?

In addition to remedies for copyright infringement, remedies are available for violation of DMCA Section 1202(b) for those who qualify for, and avail themselves of, its protections against removal or alteration of CMI, as noted more fully in our comments in response to question 26. There are, however, strong arguments that these existing claims need to be updated and augmented to address the problem of unauthorized use in training models. For example, with respect to infringement claims, a rightsholder may have a predicate issue of providing sufficient proof to evidence that its particular work was utilized for training purposes. The heightened 'plausibility' pleading requirements put into place with the Supreme Court's decisions in *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) and *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 556 (2007) could therefore prevent claims (that could have been ultimately meritorious) from surviving a motion to dismiss and reaching discovery. And Section 1202(b) claims require proof that removal or alteration of the CMI was intentionally/knowingly carried out by the AI modelers, *see* 17 U.S.C. § 1202(b)(1), and that it was done with knowledge or reasonable grounds to know that it would "induce, enable, facilitate, or conceal an infringement." 17 U.S.C. § 1202(b). The unprecedented challenges of assessing permissions conveyed in CMI for AI data mining and model training demand a change to these requirements. As detailed further in the response to Question 26, Digimarc would support the amendment of Section 1202(b) to strengthen the protections available to prevent the removal or alteration of CMI (particularly given the threat to rightsholders from the unauthorized use of content in the training of AI models) by reducing the state-of-mind evidence required to prevail on such a claim, or it would support the implementation of a separate cause of action to better secure compliance with rightsholder objections to use of their content in AI training.

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?

Developers of AI models should be required to collect, retain, and disclose records regarding the materials used to train their models. It is not possible to ascertain from the output which materials were used to train an AI model. Among other things, this could make pursuit of infringement claims against the authorized use of copyrighted content in such training difficult since the *Twombly*/*Iqbal* pleading standards seemingly require plausible identification of works used in such training, which could prevent such claims from surviving a motion to dismiss and reaching discovery. Thus, there is a need for transparency in the materials used to train AI models. Even with such transparency, it will remain a challenge to ascertain whether use of these works have been authorized for training. With a scheme for conveying rights as described in our Comments to question 9 and subparts, AI model providers and their suppliers have the means to identify these rights with particularity. Thus, in addition to retaining records of training materials, developers of AI models and their training sets should also be required to scan for CMI, including digital watermarks, as further described in our Comments to question 26.

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

When AI model providers and their suppliers detect CMI conveyed in content, they should utilize it to determine whether they have permission from the rightsholder. Digital watermarks enable the rightsholder to link content to a registry that conveys rights information including conditions for use in AI model training. When requested by the rightsholder, the AI model provider should provide notice of which works have been used in model training. This is facilitated by the registry, which enables the AI model provider to provide this notice to the rightsholder.

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?

Generative AI systems are exposed to potential liability under 17 U.S.C. 1202(b), where they do not respect CMI. Rights holders can avail themselves of the protections of Section 1202(b) by conveying CMI with their works. Certain forms of CMI, such as digital watermarks[6], can then enable generative AI systems to identify content not allowed for training in the acquisition of training materials and in the training itself.

It is a violation of DMCA Section 1202(b) to:

> Without the authority of the copyright owner or the law:
>
> **(1)** intentionally remove or alter any copyright management information,
>
> **or**
>
> **(3)** distribute, import for distribution, or publicly perform works, copies of works, or phonorecords, knowing that copyright management information has been removed or altered without authority of the copyright owner or the law,
>
> knowing, or, with respect to civil remedies under section 1203, having reasonable grounds to know, that it will induce, enable, facilitate, or conceal an infringement of any right under this title.

The training of an AI model without regard for the CMI embedded in scraped content likely leads to its removal, or at a minimum, its alteration. After all, the AI model constitutes an amalgamation of the content in the training set and a typical training process will, at a minimum, modify the CMI contained within materials in the training set. Further, since the output of generative AI systems is a derivative of this model that amalgamates the training content, it is not associated with the CMI of the input works from which it is derived.

Thus, where the generative AI model provider intentionally removes or alters the CMI in acquisition or training of its model, it is exposed to liability for inducing, enabling, or concealing infringement in the scraping and training processes as well as the outputs of the system.

Though Section 1202(b) provides protections for those that convey CMI with digital watermarks, generative AI poses a myriad of challenges not fully contemplated when the DMCA was enacted a quarter-century ago. Those scraping content for training or acquiring training data from others may ignore the presence of watermarks or contend that they are unaware of it. The lack of transparency of and access to training data makes it difficult for a rightsholder to establish that watermarked content came into the possession of an AI model provider. Further, the digital watermark is not retained in the AI model or its output in a manner that can be detected.

To further strengthen the law, the providers of generative AI models and suppliers of training data sets should be required to scan for machine-readable digital watermarks prior to the use of any content in their model training. DMCA Section 1202(b) can provide protection to content owners if digital watermarks are present, but the ingestion rate of these generative AI engines is beyond what could have

[6] Digital watermarks that convey a digital code that is imperceptible during normal use but readable by computers and software are a protectible form of CMI under DMCA section 1202. See: *IQ Group v. Wiesner Publ'g, Inc.*, 409 F. Supp. 2d 587, 596-97 (D.N.J. 2006) (noting legislative history referring to "customary indicia of ownership or authorship, such as a standard and accepted digital watermark or other copyright management information"). The *IQ Group* case cites to Digimarc's watermarking as a protectible form of CMI.

been imagined when the DMCA was written. Moreover, as noted in our Comments on question 15, there is a lack of transparency and access to the training data sets used for AI model training. As such, it is not possible for rightsholders to use digital watermark tools to check for the presence of watermarks in training data. Once content is incorporated into an AI model, the watermark is altered or removed and is not conveyed in the output of the model. The providers of AI models and their training sets are the only ones in a position to scan content used for training for digital watermarks. When provided notice from the rightsholder that its content is watermarked and reasonable access to the means to detect the watermarks and the CMI associated therewith, the generative AI model providers and suppliers should be required to scan for digital watermarks.

It is important to note that there are benefits to generative AI providers and suppliers from this required scanning as well. Watermark detection will not only help them comply with copyright laws, including the Section 1202(b), and avoid associated legal exposure and forced model re-training, but also help them avoid model collapse and gain efficiencies in model tagging. Digital watermarks provide persistent links to reliable metadata which improves the acquisition and maintenance of accurate descriptions of training data.

The massive scale of automated AI data mining, training and output and lack of transparency into training data necessitates updates to the DMCA. The double-scienter requirement of Section 1202(b) demands plaintiffs show that defendants (i) removed or altered CMI intentionally (under subsection (b)(1)) or distributed CMI knowing that the CMI was altered or removed without authority (under subsection (b)(2)); and that defendants (ii) knew or had "reasonable grounds" to know that such actions "will induce, enable, facilitate or conceal an infringement." *See* 17 U.S.C. § 1202(b). Section 1202(b) therefore requires proof of intentional removal or alteration of CMI (or distribution of works knowing the CMI has been removed or altered) *and also* actual or constructive knowledge that such removal or alteration will lead likely aid in further infringement. Among other things, to state a cognizable claim, a plaintiff must demonstrate that either a work "came into Defendant's possession with CMI attached, and Defendant intentionally and improperly removed it" or a work "came into Defendant's possession without CMI attached, but Defendant knew that CMI had been improperly removed, and Defendant used the [work] anyway." *Merideth v. Chi. Trib. Co.*, 2014 U.S. Dist. LEXIS 2346, at *7–8 (N.D. Ill. Jan. 9, 2014). A plaintiff may be unable to show how or in what form a work came into the defendant's possession in the first place and have sufficient evidence showing that the removal/alteration was done with the actual or constructive knowledge that it would facilitate infringement. For example, one court has even held that an erroneous belief about the copyright status of an image could preclude a finding of the knowledge required to state a claim under section 1202. *See, e.g.*, *Schiffer Publ'g, Ltd., v. Chronicle Books, LLC*, 2004 U.S. Dist. LEXIS 23052, at *45 (E.D. Pa. Nov. 12, 2004) (holding that a plaintiff's subjective belief that the disputed work was not under copyright protection precluded imposition of liability under Section 1202).

A mandate that generative AI models and suppliers of training data sets scan for machine-readable digital watermarks can, quite critically, help address this first scienter requirement. Without such a mandate, generative AI providers can arguably bury their heads in the sand and pretend that digital watermarks do not exist, thereby disingenuously claiming that any removal or alteration of copyright management system embedded in the digital watermarks could not have been intentional or knowing. Alternatively (or additionally), the *mens rea* standard for intentional alteration or removal could be amended or clarified to include knowingly creating AI systems that will or are likely to remove or alter CMI. This clarification is pressing given that much of the removal or alteration of CMI in the course of AI training will be large-scale and automated. Furthermore, such a clarification would address some extant case law that has found that removal of CMI that occurs as a purportedly "unintended side effect" of a technology process, such as scraping images from the internet, does not, by itself, meet the scienter requirement of Section

1202(b) *See, e.g.*, *Kelly v. Arriba Soft Corp*., 77 F. Supp.2d 1116, 1122 (C.D. Cal. 1999), *rev'd on other grounds by* 336 F.3d 811 (9th Cir. 2003) (granting summary judgment to the defendant on the plaintiff's section 1202 claims on the grounds that, inter alia, "[the p]laintiff has not offered any evidence showing [the d]efendant's actions were intentional, rather than merely an unintended side effect" and rejecting Section 1202 liability arising from search engine's crawler failure to "include [CMI] when it indexed the images"); *Logan v. Meta Platforms, Inc.*, 636 F. Supp. 3d 1052, 1064 (N.D. Cal. 2022) ("Unlike editing a plaintiff's watermark out of a photo, automatically omitting CMI by embedding a photo out of the full context of the webpage where the CMI is found cannot itself plead intentionality as required by the DMCA."). While CMI in metadata can be stripped in normal processing, CMI conveyed via robust digital watermarks in content cannot be rendered unreadable without malicious processing. Thus, the removal or alteration of such watermarks in creating an AI model and providing AI output should be prohibited.

With respect to the second scienter requirement relating to actual or constructive knowledge of inducing, enabling, facilitating or concealing an infringement, requiring AI models and suppliers to scan for digital watermarks can help support an inference of an intent to conceal infringement particularly since, while one may disagree whether some output from an AI system may constitute fair use, there is little doubt that at least some of the output created by AI models will be necessarily infringing (*e.g.*, output responding to a prompt such as "write a musical theater version of Star Wars" on a model trained on *Star Wars*). And, of course, the very scraping and training process for AI models can involve unauthorized exploitation that constitutes an infringement also cognizable (beyond just infringing output) under Section 1202(b). *See, e.g., Shihab v. Complex Media, Inc.*, 2022 U.S. Dist. LEXIS 148034, at *11 (S.D.N.Y. Aug. 17, 2022) ("[T]he Court concludes that [plaintiff] has plausibly alleged that [defendant] knew or had reasonable grounds to know that its removal of the CMI from the Photographs would induce, enable, facilitate or conceal an infringement—specifically, its own alleged infringement of [plaintiff's] copyrights to the Photographs."). Nevertheless, some brightline rules on activities that necessarily constitute an intent to induce, enable, facilitate or conceal an infringement could be an important part of ensuring that Section 1202(b) can be responsive to the needs of rightsholders in the context of generative AI. For example, according to at least one court, even intentionally cropping out a copyright notice from an image may, by itself, be insufficient to meet the intent-to-facilitate requirement. *See, e.g.*, *William Wade Waller Co. v. Nexstar Broad., Inc.*, 2011 U.S. Dist. LEXIS 72803, at *12–13 (E.D. Ark. July 6, 2011) (dismissing 1202 claim at summary judgment stage where there was no evidence that defendant's intentional cropping of copyright notice out of picture was done to "induce, enable, facilitate, or conceal infringement"). Thus, a defendant could always claim (even disingenuously) that removing a copyright notice was done for purely aesthetic or formatting reasons and potentially evade Section 1202(b) liability. These excuses are not justified for CMI conveyed via digital watermarks because they do not impact aesthetics and survive formatting. Precluding such defenses in the generative AI context would assist rightsholders ensure respect for their CMI at the advent of this era where the integrity of CMI has taken on a profound new importance.

There are, of course, myriad new challenges posed by AI to extant copyright laws, and the impact of AI on Section 1202(b) is no exception. The particular ways in generative AI companies train models and generate AI output will doubtlessly create factual circumstances not yet tested by the courts. As a consequence, application of Section 1202(b) to generative AI involves a great deal of unpredictability, and it is uncertain exactly how the jurisprudence (and the technology) will evolve. Nevertheless, we appreciate the Copyright Office's efforts to get ahead of these issues and believe that Section 1202(b) can serve as a powerful tool for ensuring respect for content owners' rights, particularly with the amendments/clarifications we have advanced here. The intent of the DMCA was to protect copyrights in

the digital world, and the recommended updates would enable the law to fulfill this intent in view of the vast technological changes over the last 25 years.

> *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?*

Governments have proposed the use of digital watermarks to label AI generated material. We applaud the governments that are having these conversations, as well as the technology companies that have indicated initial commitments to watermark AI generated output. However, the mere act watermarking of AI generated output will not provide the reliable indicator of generative AI content that sound public policy should require, and further leaves the broader questions of intellectual property rights unanswered.

A policy where watermarking occurs only at the output stage of generative AI models includes the following shortcomings:

1. So long as some AI model output is not watermarked, the lack of a watermark is not a reliable indicator that content has not been generated by AI. This is especially true given the prevalence of open-source AI models.
2. Labeling content as 'AI generated' fails to address the broader need to reliably convey CMI, including for content partially generated by AI.
3. The focus of labeling AI output has been on authenticity, yet only labeling AI output as AI generated does not provide a useful indicator of whether content lacks trust, is misleading, or has been tampered with. Generative AI is an incredible productivity tool and will be used by legitimate content creators. Conversely, inauthentic digital content has been around longer than Generative AI, as anyone who has heard the term 'deepfake' is aware.
4. Context matters both when determining whether content may be used or copied, as well as determining its authenticity. Labeling content as AI generated does not address the issue of whether content is authorized for a given use or authentic. For example, content created for one purpose may be fairly characterized as authorized and authentic in one context, and, at the time very same, not authorized or authentic in contexts outside of the one for which it was intended. For example, an image is authentic when used in an authorized channel to promote a product, yet that same image is misleading and lacks authenticity when used in an online marketplace to sell fake goods.

We share the urgency to reliably identify content generated by AI models. In addition, we seek to address the broader concern of protecting rights holders by reliably conveying rights information, whether content is generated by AI in part, wholly, or not at all. Thus, the focus should be on applying digital watermarks at the device level for a few reasons:

1. Adoption and compliance can happen quicker. This is a smaller ecosystem of players, and many of these companies have already made voluntary commitments to label AI output.
2. With a high overlap of companies that have already made voluntary commitments, compliance will be higher.
3. This expands the opportunity for content creators to supply CMI and associate source and history (provenance) with media content when it is created and not limited solely to AI model output.

Device-level watermarking is most effective at achieving ubiquitous and reliable labeling, close to the source of content creation.

4. Devices can also capture post image creation provenance, necessary to address the fact that context matters.
5. Watermark detection is also needed for the labeling to be effective, and detection is best done on devices where content is created and consumed.

In sum, regulatory efforts to mandate watermarking of AI output should facilitate the development of a broader solution to associate CMI, provenance and authentication information with content. Regulators are developing rules for what constitutes AI output, requiring labeling with digital watermarks. The need for labeling content transcends just labeling AI and requires a reliable way to label content with CMI, provenance and authenticity information. Rather than being confined to labeling AI output, digital watermarks should be used to associate CMI, provenance, and authentication information with content.

To this end, work is now underway within C2PA to use digital watermarks to bind content to content credentials. In the C2PA standard, the content credentials provide provenance and authenticity in metadata called a manifest. Digital watermarks provide a way to reliably bind the manifest to content as well as provide CMI. This approach enables content creators to assert their rights in content as well as convey provenance and information to detect tampering.

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

AI application providers need to provide the capability to identify AI-generated content and enable users to label their content appropriately. Regulators are in the process of formulating requirements for watermarking AI generated content, and this will establish a baseline for labeling requirements for AI output. Users will leverage this output to create content, including content that is not entirely AI generated. Thus, users need the capability to label content with CMI, provenance and authentication information. The use of digital watermarks to associate content with metadata that addresses all of these needs would represent a preferred scheme. As described above, digital watermarks can bind content to its metadata, such as a C2PA manifest, as well as CMI. This enables reliable and efficient detection of rights and provenance information. For rights management and authenticity, this is a fundamentally more effective approach than a simple indicator of AI output because it enables more granular rights assertion and provenance information tied to content, including content generated partially by AI and human creative input.

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

Technical solutions are available for labeling content using digital watermarking. As noted, a preferred way to facilitate compliance with labeling requirements as well as convey CMI to manage rights is to use digital watermarks to reliably link content with its CMI and provenance metadata.

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 DMCA Section 1202(b) provides a remedy for alteration or removal of CMI. As noted, AI models give rise to a myriad of new challenges that necessitate updates to the DMCA to facilitate detection and adherence to the rights asserted in CMI as described in our comments on question 26.

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?

No tool has proven effective in identifying AI generated output if it was not labeled as such by its creator. Digital watermarking tools exist to identify AI-generated material and are most effective when used to bind content to CMI and provenance information. C2PA has developed a standard for attaching provenance data to content, and this provenance data may provide a content credential that indicates the AI model used and version to create content. Since this provenance data is attached as metadata which can be stripped or swapped, a digital watermark should be used to reliably bind the content credential to the content.

Digimarc thanks the United States Copyright Office for its consideration of these comments.

Sincerely,

Joel Meyer

Chief Legal Officer

Digimarc Corporation