DIGIMARC

The need for standardization and ubiquity in the use of digital watermarks

Riley McCormack, CEO
Digimarc
November 21, 2023

Leader Schumer, Senator Rounds, Senator Heinrich, and Senator Young:

I am grateful and honored to have the opportunity to meet with you on the pressing issues of transparency, explainability, intellectual property, and copyright as they pertain to the advent of artificial intelligence (AI). I am the CEO of Digimarc, a product digitization company that pioneered digital watermarking technology almost thirty years ago. The original application of our technology was to protect the copyrights of creators and inventors at the advent of another technology revolution: the internet. Since then, our technology has been deployed at a global scale many times over, by both the public and private sectors, not only to help ensure that the rights of creators and inventors are protected, but also to ensure the authenticity and trustworthiness of physical and digital assets worldwide. It is because of our work in this domain that our technology has been trusted by a consortium of the world's central banks to protect global currencies for nearly 25 years.

Given our long history with digital watermarking, we believe we have insight into how this powerful technology could be applied to address the pressing issues upon which this Forum is focused. We also believe it is important to highlight that not every technology currently being given the moniker of "digital watermarking" qualifies as such, nor are all digital watermarks created equal. True and reliable digital watermarking, however, is a tested and effective technology that can and should be used to help create a safer, fairer, and more authentic internet. Digital watermarks can communicate copyright information and provide transparency by helping consumers understand key information like the provenance of, or the use of AI in, the digital content they consume.

Standards to strengthen accountability and boost innovation

It is exactly because of what could be enabled by the application of true and reliable digital watermarking, combined with the inconsistent or incorrect use of this term, that standards that enable long-term integrity in digital watermarking must be created and adhered to ubiquitously across the US market and beyond.

Digimarc lauds the Administration's recent *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* as it creates a role for the National Institute of Standards and Technology (NIST) to develop these needed "fit-for-purpose" technical standards to address trust model gaps for tools like digital watermarks. Doing so is essential to elevating safety, security and trust in digital content while also enabling mutual accountability.

A sign in the front window of your house reading "Beware of Dog" should not be characterized as a lock. In addition, a hook lock attached to a screen door does not provide the same security as a multi-pronged deadbolt attached to a steel door despite both being characterized as locks. Today, there are numerous technologies that claim or are considered to be digital watermarks when they are not, and in addition, the technical variance between different versions of digital watermarking is sizable.

We need agreement on what constitutes digital watermarking that can only come from the creation of a standard, and then we need adherence to that standard. The utility of any technology comes from its ability to achieve critical tasks, not in just the application of a name.

How Digimarc defines digital watermarks

Digital watermarking is the science of hiding information about an item in the item itself. Images, audio, video, and documents are among the type of digital assets that are currently digitally watermarked at global scale. Embedding is the act of hiding the information in an asset and the process of discovering that information is referred to as detection. Often, this embedded information consists of a unique identifier, and that identifier is used to communicate content provenance, authenticity, and copyright information about a digital asset in a way that is both secure and inextricably linked to the asset itself.

Digimarc believes trustworthy and reliable digital watermarks have five specific characteristics, all of which must be present to ensure system integrity:

1. **Covert and Machine Readable:** A digital watermark is meant to be read by machines, not humans. As such, in addition to the requirement of indeed being detectable and readable by machines, a digital watermark can also be (and almost always is) invisible and/or inaudible.
2. **Immutable:** Once a digital watermark is applied to a digital asset, it becomes part of the digital asset itself, and thus the information cannot be changed. This also enables a digital watermark to transit all formats and distribution channels, ensuring its utility regardless of workflow.
3. **Ubiquitous:** Because a digital watermark is covert, it can cover the totality of a digital asset, thereby making it more difficult to damage, manipulate, or remove the information.
4. **Redundant:** Since a digital watermark can cover the totality of a digital asset, in addition to how the digital watermark itself is constructed, the digital asset can suffer significant damage or manipulation without impacting the digital watermark's ability to be detected.
5. **Secure:** The information contained in a digital watermark should only be shared in encrypted form which, when combined with other properties listed above, results in a robust data carrier that cannot be trivially severed from an asset. This also means digital watermarks should not be created using open-source technology as this introduces system risk caused by bad actors.

For a digital watermark to be effective it requires two critical steps: embedding the digital watermark within the asset and detecting that digital watermark (and the information attached). This is logical, because for any form of communication to have value, the message must be received, not just sent. Without both steps—embedding and detection—a digital watermark is not effective, and thus has no value.

Manifests and metadata are not digital watermarks

Information about a digital asset, such as provenance and ownership, is referred to as metadata. Sometimes this metadata is stored in a format called a manifest. This information is critical to determining the ownership or authenticity of digital content, but it is not a digital watermark.

The term "digital watermarking" is sometimes erroneously used when discussing metadata and manifests, however, and this confusion has significant implications for system security. As an example, metadata attached to an image file differs from digital watermarking in that the image file is not immutable and the metadata contained therein is not ubiquitous or redundant. Moreover, metadata is noticeably insecure as it can be (and often is) easily removed from files, severing it from the content to which it was originally tied. Then, new (and incorrect or nefarious) metadata can be inserted instead. The ease with which metadata can be removed or replaced is not lessened by any cryptographic techniques used to sign the metadata.

However, when metadata, manifests, and digital watermarking are used in concert, they act like a layered security system, making the information tethered to content that much harder to sever, corrupt, or remove—

just as adding a home security system and a guard dog, on top of a deadbolted steel door, would make a home more secure.

Digital watermarks are the most critical security layer of this trifecta, as without a digital watermark the manifest and metadata are not securely and inextricably linked to the content itself. Manifests and metadata need to be adhered to content through trustworthy digital watermarks to ensure the information they contain is both present and reliable.

Digitally watermarking GenAI content is necessary but not sufficient

The recent commitments made by large technology companies to digitally watermark content created by their generative AI engines (synthetic content) is a necessary step toward creating a more transparent digital ecosystem, so long as the digital watermarking technology used in that process meets certain baseline specifications as developed and determined by a standard-setting body like NIST. We believe, however, that additional steps must also be taken to create a truly transparent digital ecosystem.

First, nefarious actors will always have access to non-compliant GenAI engines. In addition, no single-layer security system is infallible, especially when the rewards for breaking that system are high. No trustworthy and reliable system of authenticity has ever been built upon solely marking non-authentic content as such, and the fact that many large technology companies have agreed to digitally watermark the output of their GenAI engines should not lull us into a false sense of security that the problem is solved. Instead, these commitments should be viewed as a powerful first step in providing some coverage to one side of the trust and transparency coin. To provide true value to that coin, however, the tools to digitally watermark non-synthetic content must be made easily accessible at all points of content creation and distribution so that non-synthetic content creators have the option to digitally watermark their digital assets, too. Doing so is not only the best way to create a true ecosystem of trust and transparency, but it would also increase the difficulty for bad actors intent on breaking the system.

Second, software to detect the digital watermarks used in synthetic and non-synthetic content must be made readily available. As mentioned previously, for any form of communication to have value, the message must be received, not just sent. The focus on requiring GenAI engines to digitally watermark their output has no value if a corresponding focus on the methods and means to make sure those digital watermarks are detected is absent. This same detection software should be able to detect the digital watermarks on non-synthetic content as well, thus providing value to both sides of the required two-sided trust and authenticity coin.

Device-level embedding and detection of digital watermarks

The fact that not every global GenAI company, social media network, app, or website will adopt digital watermarking embedding and detection creates loopholes and provides an ecosystem of false certainty—and when it comes to the integrity of a system of trust and authenticity, false certainty can cause much more damage than uncertainty.

GenAI engines and applications are proliferating which makes the ecosystem of AI more multinational and complex. The democratization of AI tools means that synthetic content will not only be produced by large tech companies, but also by individuals and smaller entities using these tools. This means that synthetic content will proliferate while only the synthetic content produced by certain GenAI companies will be marked as such, and, even then, only in certain situations.

The social media networks, apps, and websites where this content is displayed is also heavily fragmented and global in scale. Even if all synthetic content were digitally watermarked as such, and all creators of non-synthetic content had easily accessible tools to digitally watermark their content as well, these non-compliant points of content creation and consumption will jeopardize the integrity of the entire system.

To close these loopholes, Digimarc advocates that the embedding and detection of digital watermarks be done at a device level—on smartphones, computers, tablets, servers, TVs, etc. All digital content is created and consumed on a digital device, therefore the most foolproof way of achieving ubiquity for the embedding and detection of digital watermarks is at the device level.

To enact this, a device manufacturer would simply need to include digital watermarking embedding and detection software as part of the next software update it sends to all its devices (like the smartphone software updates that happen every few months). Should device manufacturers agree to include this technology in their next update, device-level digital watermarking embedding and detection technology could be ready and deployable to billions of devices within a matter of weeks.

Then, whenever new synthetic content is created and distributed, a digital watermark would be embedded in that content to ensure the communication of key information like provenance and GenAI usage. Creators of non-synthetic content would also have access to digital watermarks at the point of publication to communicate the provenance and ownership of their content. Lastly and importantly, whenever digitally watermarked content would be displayed in an app or on a browser, the digital watermark would be detected and consumers could be notified of the digital content's corresponding information, thereby helping create a more transparent digital ecosystem. On-device embedding and detection is the most secure, efficient, and ubiquitous way to deploy digital watermarks at scale, and therefore create a safer, fairer, and more transparent internet.

An on-device system of digital watermarking would also provide opportunities for content creators to digitally watermark their content as authentic and original at the moment of publication, thereby enabling them to easily claim (and protect) their copyright. By securely attaching copyright information to digital content in a machine-readable way, digital watermarking not only makes it easy for content creators to claim their copyrights, it also helps ensure that content creators' work is not inadvertently used to train GenAI models without their knowledge, consent, and compensation. These same digital watermarks could also be used by the GenAI models to help protect against model collapse and enable efficient model training by ensuring the relevant information about a digital asset is always available for training purposes.

Closing thoughts and commitment to partnership

Digimarc believes that the potential for innovation and growth offered by AI is tantamount to none, but the power of AI must be safely regulated in an effective way. If we fail to get it right, there will be serious ramifications for generations to come. AI, and the tools that are intended to safeguard it, can be misused by bad actors and thus must be governed by enforceable guardrails.

Leaders in Washington, D.C. have an immense task ahead of them in working to find the right balance for regulating an ever-evolving and revolutionary technology. Solutions that can both promote innovation and protect against misuse must be seized. We believe that the Senate AI Insight Forums and the Executive Order on AI released in October of this year are critical steps that will inform the necessary path forward.

On behalf of the team at Digimarc, let me say how grateful I am to be included in this forum, and I would like to sincerely thank Leader Schumer, Senator Rounds, Senator Heinrich, and Senator Young for the opportunity to participate. Digimarc looks forward to doing our part to help find the most effective way forward in this new frontier.