Digimarc adds copyright information to digital data

Software company Digimarc will now let copyright owners add more information to their work, which the company said will improve how AI models treat copyright in training data.

In a statement, Digimarc said its new Digimarc Validate service lets users include ownership identification in the metadata. The company said this means that when copyrighted material becomes part of a generative AI training dataset, users can point to the digital watermark with intellectual property information.

For example, an image with Digimarc Validate adds a © symbol that is machine-readable and includes information on who owns the copyright. The company said Digimarc Validate is powered by its digital watermark detection software, called SAFE, or secure, accurate, fair, and efficient, which AI companies have to buy into if they want to prevent copyrighted material with the Digimarc Validate symbol from making it to training datasets.

“Generative AI has changed the rules, and once digital assets are distributed or published, the ability to protect those valuable assets is gone,” said Digimarc president and CEO Riley McCormack.

Digimarc said much of the content in datasets scraped for AI training “is copyrighted; it’s just not digitally identified as such.” This allows generative AI models to identify which data is protected before the model ingests it for training.

Noting copyright ownership in the metadata of content sounds

great on paper, but it will only work if AI developers actively avoid copyrighted material to train models. So far, AI companies have not promised they will stay away from copyrighted material in training datasets. However, having a digital paper trail of copyright could allow creators to point out if AI developers do intentionally infringe protections.

Some AI companies, like Adobe, said they only use licensed data for training. OpenAI announced websites can block its web crawler so it doesn't take in that data for training.

Meanwhile, Microsoft has said it will take the legal heat if commercial customers using its Copilot products get sued for copyright infringement.

Some of the first few lawsuits against developers of generative AI models deal with the thorny issue of copyright infringement. Comedian Sarah Silverman and authors Christopher Golden and Richard Kadrey sued OpenAI and Meta for allegedly using their books to train GPT-4 and Llama 2. Three artists filed a lawsuit against Stable Diffusion, Midjourney, and the art website DeviantArt for allegedly infringing their copyright.

To help figure out how to approach AI and copyright issues best, the US Copyright Office opened a public comment period on August 30th to understand people's concerns.

While the White House got commitments from AI companies to develop watermarks, the focus of those watermarks is to identify AI-generated content.

"The risk content owners face from failing to add an identified to digital assets before distribution or publication goes beyond just misuse and theft," said Digimarc chief product officer Ken Sickles. "In the future, your digital assets will make ecommerce transactions more trustworthy, email more secure, and social media a safer place."

Digimarc Validate is available for commercial use, beginning at $399 a month. Enterprise customers can work with the company for pricing options.

https://www.theverge.com/2023/9/19/23879555/digimarc-copyright-watermark-generative-ai