Artistic Licence: Can Copyright Rein in Generative AI?
Muireann Bolger
February 2, 2023

In part II of a series on art and AI, Muireann Bolger hears how lawsuits against machine learning 'artists' may end with Spotify-style databases of licensed works.

When artist and illustrator Kelly McKernan discovered that her artwork had been used to train the artificial intelligence (AI) art tool, Stable Diffusion, she initially found it "surreal and exciting".

But as she later revealed in a tweet, this elation swiftly turned to "nauseating" anger—prompting a lawsuit that could have crucial implications for both AI developers and copyright owners.

After realising that her "art had been trained on without her consent or compensation", McKernan, and fellow artists Sarah Andersen and Karla Ortiz, brought a class action complaint against the tool's creator, Stability AI, earlier this month.

The case marks the start of a flurry of litigation against the developers of generative AI— systems that can learn concepts from large bodies of existing knowledge and which then use what they learn to help people create new works.

In London, Getty Images sued Stability AI for using images for training purposes without its permission, while Microsoft, Microsoft's GitHub, OpenAI and MidJourney have been accused of infringement by allegedly using copyrighted open source code to train their machine learning (ML) systems.

International repercussions

But according to lawyers and academics, it is the artists' case that has the power to reverberate across the legal sphere, and to truly divide opinion.

For Mark Nichols, senior associate at Potter Clarkson, the US case covers more territory than the Getty lawsuit, as it explicitly looks at the IP implications regarding content that has been created by an AI—even if it has used the work of others to do so.

"It asks whether the output from an AI infringes as a derivative work, whereas the Getty case just concerns the content that was an input," he explains.

As a result, its outcome could have significant repercussions internationally, he adds.

"While the law differs from country to country, we can expect the technology used in generative AI to be explored thoroughly and this will be informative and, potentially, persuasive in future disputes regardless of where that dispute is taking place."

McKernan and her co-plaintiffs argue that the derived images from Stable Diffusion will ultimately compete in the marketplace with the originals—and that buyers will be able to access the artists' works to generate new works without compensating them.

Stability AI, however, strongly rejects this stance, arguing that the trio have a fundamentally flawed vision of the purpose of generative AI.

A spokesperson from the company told *WIPR* that: "The allegations in this suit represent a misunderstanding of how generative AI technology works and the law surrounding copyright. We intend to defend ourselves and the vast potential generative AI has to expand the creative power of humanity."

Fair use and transformative works

The crux of the dispute will ultimately come down to how the US courts will apply the doctrine of fair use—the right to use a copyrighted work under certain conditions without the copyright owner's permission.

For some, Stability AI has a good case for countering the claims.

Ryan Abbott, partner at Brown, Neri, Smith & Khan, points out that the fair use rule has been successfully used as a basis to avoid liability in cases involving text and data mining.

"Fair use is a potential defence to copyright infringement based on a court's analysis of factors such as how much of a protected work is being used and the purpose of that use. It has been and it may also be a defence to use images as training data for a generative AI such as Stable Diffusion," he reflects.

This view has curried favour in academic circles: Rebecca Tushnet, professor at Harvard Law School, draws a comparison between the leverage granted to humans when coming up with 'transformative works'.

This is when a new work based on an old one is deemed transformative if it uses the source work in completely new or unexpected ways, enabling the creator to avoid accusations of infringement.

She explains: "In general, the human capacity to absorb influences and come up with new works in response is usually accepted as how creativity works and a good thing rather than a bad one, outside of cases in which there is infringing similarity between two outputs."

Teaching a computer to "create" by training it on existing works is a very similar concept, Tushnet concludes.

Mark Lemley, professor of law at Stanford Law School, has gone a step further.

In a co-authored paper, "Fair Learning" published in the *Texas Law Review*, he argued that because ML training sets are likely to contain millions of different works with thousands of different owners, there is no plausible way to simply licence all of the underlying photographs, videos, audio files, or texts for the new use.

"The vast potential of ML systems is matched only by their appetite for data. To perform, they must first learn how—generally, through a process of trial-and-error of epic proportions. And in order to create the right conditions for this learning process, engineers must begin by collecting and compiling enormous databases of exemplary tasks for machines to practise on, known as 'training sets'," he wrote.

The paper goes on to say that an ML system's use of the data "often is transformative as that term has come to be understood in copyright law", because even though it doesn't change the underlying work, "it changes the purpose for which the work is used".

A 'legally toxic' end product

Such arguments, however, are anathema to Mark Milstein, a co-founder of vAIsual, a technology company pioneering algorithms and solutions to provide licenced content to generative AI developers.

In his view, developers such as Stability AI, Midjourney and Open AI, "have opened a hornet's nest of negatives", having approached their training sets in "completely the wrong way".

"Have they gone through the proper legal channels and licensed their content? The answer is no," he says, going on to fervently reject suggestions that they lack the means to do so.

"My company, for example, has data sets for millions of images owned by people who have allowed their images to be used exclusively for the training of generative AI," he explains.

Adding that the AI developers "have crossed a Rubicon"— either by commercialising the product, or by allowing end users to commercialise their outputs—Milstein argues that their endeavours have, ultimately, been in vain.

"In a word, the fair use argument is kaput," he says disdainfully, adding that these companies have merely ended up with a "legally toxic and radioactive end product".

A vast challenge

Disruptive technologist and founder of the Innovation Foundry, John Collins echoes Milstein's insistence on the need to police generative AI— but is less convinced that the tools or regulations currently exist to do so.

"Stability AI has taken someone else's data without permission and used it without authorisation. Now, if I were to do that, for example, take someone else's data and use it in a paper for publication, I'd be called a plagiarist. Then, why shouldn't Stability AI be described accordingly?"

Yet, in his view, many unwitting rights owners now face an unenviable task—trying to navigate a labyrinthine journey in search of stolen data that will, in many cases, result in dead ends.

"When one thinks about the sheer vastness, the sheer scale of the internet, policing and finding data that has been misused will be almost impossible," he explains.

"And when enough people realise that this is happening, there will be a global challenge to current IP law— meaning that it will have to adapt to changing times and embrace what's going on in technology, and how it's potentially destroying IP support for many," he predicts.

The same rules apply

Unsurprisingly, lawyers take a more pragmatic approach to current issues, arguing that existing laws will be able to withstand the onslaught of questions posed by AI.

For example, Maurico Uribe, partner at Knobbe Martens, argues that "there is nothing unique to AI-based implementations that would suggest that existing statutory and legal doctrines would expressly allow for such unauthorised use, and otherwise inhibit a rightful IP owner from accessing the available forms of damages or equitable relief."

As Blair Jacobs, principal at McKool Smith, puts more bluntly: "Copyright violations are copyright violations, no matter how they are to occur. And I think that the same rules still apply."

In his view, copyright infringement remains "the easiest to prove violation" of all of the different forms of IP contraventions, and dismisses predictions of a potential seachange in IP law as "an overreaction".

"As technology evolves, we do see new and interesting lawsuits centred on this technological innovation and novel ways of presenting information to people. But both UK and US courts have tended to successfully and consistently apply old precedent to the new technological developments," he says.

"Despite the fact that the technology may be a little bit more complicated, we still are left with questions concerning the wrongful use of copyrighted material, regardless of how we get there—and that's what these suits seem to be addressing."

He draws parallels to the dilemma presented by the advent of music streaming more than a decade ago.

"We had to adjust to the fact that we had all of this music readily available on the internet, and you could for your own personal use, download it and listen to it. At first there weren't even subscription models, but

these did come—largely because of copyright, and the fact that it was recognised that the creators of the copyrighted material had to be reimbursed for what they had contributed.

"I think the same exact thing will happen with regard to AI-generated images."

Milstein also concludes that copyright remains the unsurpassed remedy—and the repositories of licensed content data for training purposes need to become the norm.

"I think generative AI is the future of visual media. And it will allow for a much greater degree of creativity—as long as everybody respects copyrights and trademarks, then we have great things to look forward to."

"This is fantastic new technology"—he concludes—"but let's get it right".