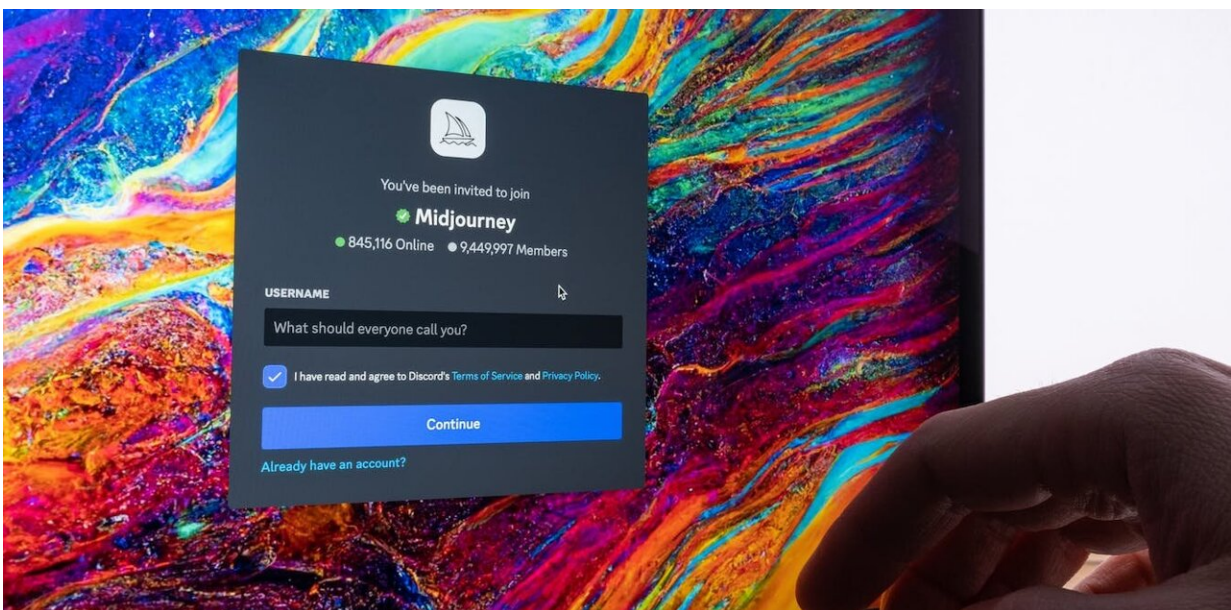


Is there a way to pay content creators whose work is used to train AI? Yes, but it's not foolproof

February 28 2023, by Brendan Paul Murphy



Credit: Shutterstock

Is imitation the sincerest form of flattery, or theft? Perhaps it comes down to the imitator.

Text-to-image artificial [intelligence systems](#) such as DALL-E 2, Midjourney and Stable Diffusion are trained on huge amounts of image data from the web. As a result, they often generate outputs that resemble

real artists' work and style.

It's safe to say artists [aren't impressed](#). To further complicate things, although [intellectual property law](#) guards against the misappropriation of individual works of art, this doesn't extend to emulating a person's style.

It's becoming difficult for artists to promote their work online without contributing infinitesimally to the creative capacity of generative AI. Many are now asking if it's possible to compensate creatives whose art is used in this way.

One approach from photo licensing service Shutterstock goes some way towards addressing the issue.

Old contributor model, meet computer vision

Media content licensing services such as Shutterstock take contributions from photographers and artists and make them available for third parties to license.

In these cases, the commercial interests of licensor, licensee and creative are straightforward. Customers pay to license an image, and a portion of this payment (in Shutterstock's [case](#) 15%–40%) goes to the creative who provided the [intellectual property](#).

Issues of intellectual property are cut and dried: if somebody uses a Shutterstock image without a license, or for a purpose outside its terms, it's a clear breach of the photographer's or artist's rights.

However, Shutterstock's terms of service also allow it to pursue a new way to generate income from intellectual property. Its current contributors' site has a large focus on [computer vision](#), which it defines as: "a scientific discipline that seeks to develop techniques to help

computers 'see' and understand the content of digital images such as photographs and videos."



The LLM process is a bit like an impartial art student learning about techniques and genres by wandering through a gallery of millions of captioned paintings. Can we say any individual painting added more to their generalised knowledge? Probably not. Credit: Shutterstock AI

Computer vision isn't new. Have you ever told a website you're not a robot and identified some warped text or pictures of bicycles? If so, you have been [actively training AI-run computer vision](#) algorithms.

Now, computer vision is allowing Shutterstock to [create](#) what it calls an "ethically sourced, totally clean, and extremely inclusive" [AI image generator](#).

What makes Shutterstock's approach 'ethical'?

An immense amount of work goes into classifying millions of images to train the large language models used by AI image generators. But services such as Shutterstock are uniquely positioned to do this.

Shutterstock has access to high-quality images from some [two million contributors](#), all of which are described in some level of detail. It's the perfect recipe for training a large language model.

These models are essentially vast multidimensional neural networks. The network is fed training data, which it uses to create data points that combine visual and conceptual information. The more information there is, the more data points the network can create and link up.

This distinction between a collection of images and a constellation of abstract [data points](#) lies at the heart of the issue of compensating creatives whose work is used to train generative AI.

Even in the case where a system has learnt to associate a very specific image [with a label](#), there's no meaningful way to trace a clear line from that training image to the outputs. We can't really see what the systems measure or how they "understand" the concepts they learn.

Shutterstock's solution is to compensate every contributor whose work is

[made available](#) to a commercial partner for computer vision training. It describes the approach on its site:

"We have established a Shutterstock Contributor Fund, which will directly compensate Shutterstock contributors if their IP was used in the development of AI-generative models, like the OpenAI model, through licensing of data from Shutterstock's library. Additionally, Shutterstock will continue to compensate contributors for the future licensing of AI-generated content through the Shutterstock AI content generation tool."



The Have I Been Trained? results turn up a CC-licensed photo I uploaded to Flickr about a decade ago. Author provided

Problem solved?

The amount that goes into the Shutterstock Contributor Fund will be proportional to the value of the dataset deal Shutterstock makes. But, of course, the fund will be split among a large proportion of Shutterstock's

[contributors](#).

Whatever equation Shutterstock develops to determine the fund's size, it's worth remembering that any compensation isn't the same as *fair* compensation. Shutterstock's model sets the stage for new debates about value and fairness.

Arguably the most important debates will focus on the amount of specific individuals' contributions to the "knowledge" gleaned by a trained neural network. But there isn't (and may never be) a way to accurately measure this.

No picture-perfect solution

There are, of course, many other user-contributed media libraries on the internet. For now, Shutterstock is the most open about its dealings with computer vision projects, and its terms of use are the most direct in addressing the ethical issues.

Another big AI player, Stable Diffusion, uses an open source image database called [LAION-5B](#) for training. Content creators can use a service called [Have I Been Trained?](#) to check if their work was included in the dataset, and opt out of it (but this will only be reflected in future versions of Stable Diffusion).

One of my popular CC-licensed photographs of a young girl reading shows up in the database several times. But I don't mind, so I've chosen not to opt out.

Shutterstock [has promised](#) to give contributors a choice to opt out of future dataset deals.

Its terms make it the first business of its type to address the ethics of

providing contributors' works for training generative AI ([and other](#) computer-vision-related uses). It offers what's perhaps the simplest solution yet to a highly fraught dilemma.

Time will tell if contributors themselves consider this approach fair. Intellectual property law may also evolve to help establish contributors' rights, so it could be Shutterstock is trying to get ahead of the curve.

Either way, we can expect more give and take before everyone is happy.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Is there a way to pay content creators whose work is used to train AI? Yes, but it's not foolproof (2023, February 28) retrieved 23 September 2023 from <https://techxplore.com/news/2023-02-pay-content-creators-ai-foolproof.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.