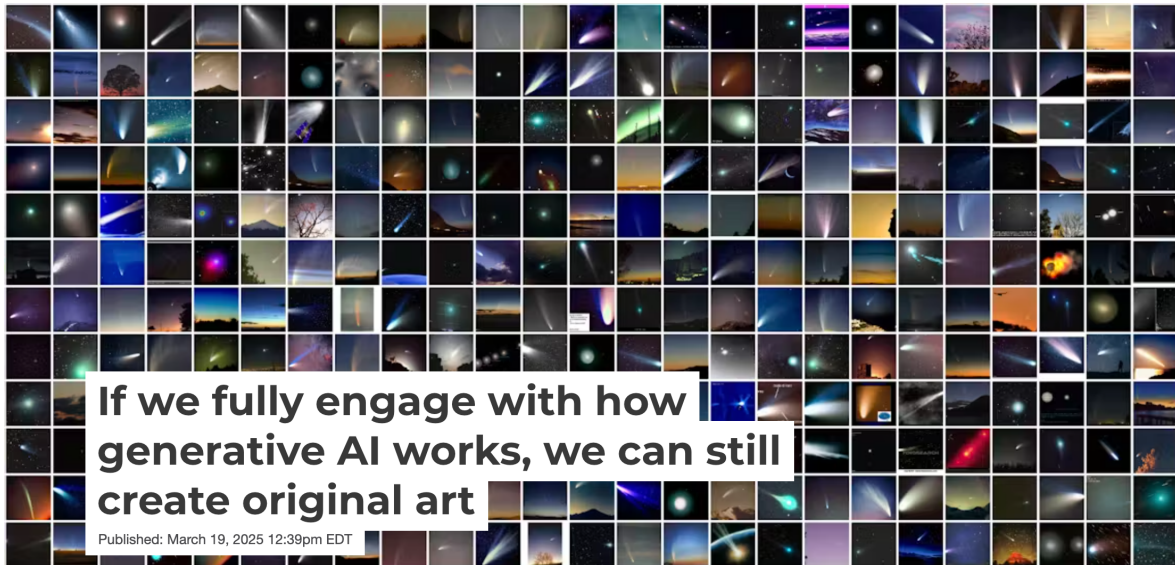# THE C○NVERSATION

Academic rigor, journalistic flair

Downey, Anthony, "If we fully engage with how generative AI works, we can still create original art," *The Conversation*, March 19, 2025

If we fully engage with how generative AI works, we can still create original art

Published: March 19, 2025 12:39pm EDT

Even before the recent protest by a group of well-known musicians at the UK government's plans to allow AI companies to use copyright-protected work for training, disquiet around artists' rights was already growing.

In early February, an open letter from artists around the world called on Christie's auction house to cancel a sale of art created with the assistance of generative AI (GenAI). This is a form of artificial intelligence that creates content – including text, images, or music – based on the patterns learned from colossal data sets.

Without giving specific examples, the letter suggested that many of the works included in the sale, which was entitled "Augmented Intelligence" were "known to be trained on copyrighted work without a licence" and suggested that such sales further "incentivises AI companies' mass theft of human artists' work."

If we think about Dall-E, Midjourney, and Stable Diffusion, all of which use text prompts to generate images and are trained on data sets harvested from online sources, the letter raised significant issues about the nature of artistic creativity and how the legal concept of "fair use" and originality is applied in such cases.

These are complex debates, encompassing perennial misgivings about machine automation, intellectual property (IP), and the cherished ideal that ingenuity and originality remain the sole preserve of humanity.

How to think from within GenAI

The impact of AI on the creative industries has become a major issue in the UK and elsewhere, so much so that we are faced with an existential question: how do we understand the evolving impact of AI on human creativity today?

The scope of this enquiry reveals a simple fact: we need to develop more accessible and inclusive ways to think from within AI image processing models. This is exactly what my latest research, produced in collaboration with the acclaimed artist and photographer Trevor Paglen, proposes.
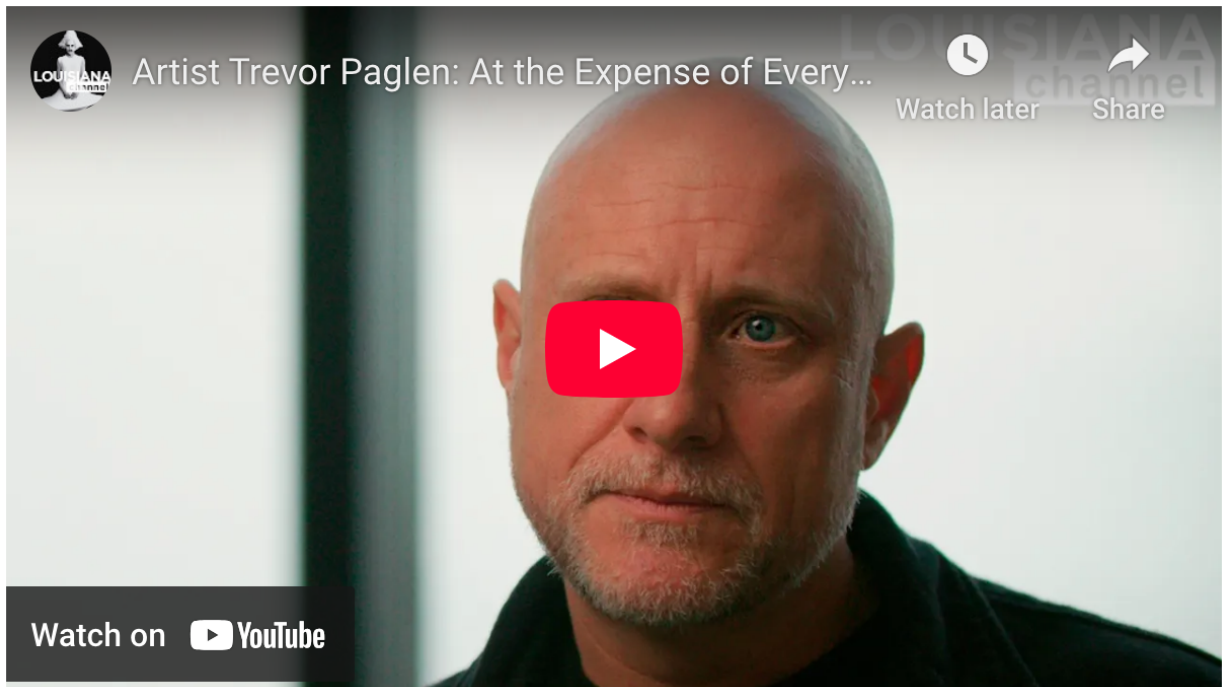
How, this research asks, do we better understand the mechanisms behind the collation and labelling of the data sets that are used to train AI? And how, in turn, can we create new ways for understanding the extent to which AI image-production models inform our experience the world?

It is, I argue, through the development of interdisciplinary research methods that draw upon the arts and humanities that we can critically engage with these concerns.

Although the open letter addressed to Christie's alluded to these topics, it did not, perhaps unsurprisingly, observe the degree to which some of the more prominent artists in the Augmented Intelligence sale had actively engaged in providing visual methods and insights into how GenAI functions.

It is notable that Holly Herndon and Mat Dryhurst's work xhairymutantx scrutinises how the data sets used in AI models of image production both define and transform images. For example, if you type the word "Holly Herndon" into Midjourney, it will produce images that are based on data sets derived from Herndon's online presence.

To draw attention to, and simultaneously disrupt, this process, the artists generated their own data sets of images and labelled them "Holly Herndon". The images in these data sets had been previously manipulated to emphasise

certain qualities associated with Herndon (her red hair, for example). Once fed back into the AI image processing model, the ensuing images of "Holly Herndon" became evermore outlandish and exaggerated.

This clearly shows that AI image processing is a highly inconsistent and selective procedure that can be manipulated with ease.

If we consider how models of AI image processing are used in facial recognition and drone technologies – often with fatal consequences – this is an urgent concern.

Reflecting upon aerial photography in his work Machine Hallucinations – ISS Dreams, artist and data visualisation pioneer Refik Anandol used a data set of 1.2 million images collated by the International Space Station (ISS). Alongside other satellite images of Earth, he produced an AI-generated composition.

Employing generative adversarial networks (GANs) – an AI model that trains neural networks to recognise, classify and, crucially, generate new images – Anandol effectively produced a unique landscape that changes over time and never seems to repeat itself.

Rainbow (Corpus: Omens and Portents), 2017. Trevor Paglen Studio, 2025, Author provided (no reuse)

In both these examples, artists are not simply engaging in either "mass theft" or using AI models that have been trained on large data sets to mechanically produce images. They are explicitly drawing attention to how the data sets used to train AI can be both strategically engineered and actively disrupted.

In our recent book (to which I contributed as editor and author), Trevor Paglen, whose work was not in the Christie's sale, reveals how data sets regularly produce disquieting, hallucinatory allegories of our world.

Given that GANs are trained on specific data sets and do not experience the world as such, they often produce hallucinatory and uncanny versions of it. Although often considered to be a fault or a glitch in the system, the event of hallucination, as Paglen demonstrates, is nevertheless central to GenAI.

In images such as Rainbow, which was produced using a data set created and labelled by Paglen, we see a ghostly image of our world that discloses the inner, latent mechanics of image production in GANs.

Paglen's practice, alongside that of Dryhurst, Herndon and Anandol, defines a clear distinction between those artists who casually use AI to generate yet more images and those who critically investigate the operative logic of AI. The latter approach is precisely what is needed when it comes to thinking through GenAI and rendering it more accountable as a technology that has evolved to define significant aspects of our lives.

If we allow that the internal workings of AI are opaque to users and programmers alike, it is all the more crucial that we explore how art practices – and the humanities more broadly – can encourage us to think from within these unaccountable systems. In doing so we could significantly improve levels of understanding and engagement with a technology that is defining the future and our relationship to it.