Generative AI is here to stay, but so are artists

Generative Artificial Intelligence is the next commercial frontier for AI. What appeared as a surreal meme generator earlier this year has turned into a goldrush with the success of companies like Stability AI. Max Lunn digs into how they work, their use cases, the issues plaguing them – and why long-termism has clouded the artist debate.

Ever imagined what Soviet Afrofuturism looks like? Me neither, but you don't have to imagine: DALL·E 2 has done it for you (see below). The digital visual landscape has been changed in a matter of months by three text-to-image AI image generators: DALL·E 2, Stable Diffusion and Midjourney. Others such as Nightcafe are also gaining in popularity. In essence, these allow those with access to generate totally new images, based on human inputted natural language prompts.

The results have been startling: the generating models can instantly create novel photorealistic images that accurately follow complex optical phenomena or mimic idiosyncratic artistic styles. They can genre-mix, time travel and predict the future in a manner that suggests a boundless, wild imagination. Remarkably, some of these image generators are open source: anyone can access DALL·E 2 for free, plugging in their own absurdities and leaving with their eery creations.

Although these three generators seemingly arrived out of nowhere, they do

have predecessors. You may recall hyper-real images of faces starting to appear a few years ago – faces that in fact didn't belong to anyone and were mere digital fictions. These arrived courtesy of 'generative adversarial networks' (GANs). Since GANs were introduced in 2014 by Google Researcher Ian Goodfellow, the tech has been widely adopted in image generation and transfer. Check out *thispersondoesnotexist.com*.

Given DALL·E 2's and others use for artistic purposes, it is worth noting the AI project '*The Next Rembrandt*'. This was a 3D printed painting based solely on data collected from Rembrandt's oeuvre, and made using deep learning algorithms and facial recognition techniques. The painting is hugely convincing, appearing as a kind of archetypal Rembrandt, and full of accurately rendered emblematic features. As one of the people involved said: 'we're using a lot of data to improve business life, but we haven't been using data in a way that touches the human soul'.

However, these models could not conjure up images from a simple text prompt: they were not nearly as accessible, or instant. Their predecessor actually came about in 2015 from automated image captioning – a mechanical version of the tedious 'reCAPTCHA' process we all are forced to do to prove our humanity. Machine learning algorithms had 'learnt' to caption images with natural language, putting them into descriptions. A group of researches from Toronto were inspired to *try this process in reverse*. One of their early successes was to create a blurry image of a green school bus, demonstrating it was possible to surpass a simple google image search equivalent, and instead invent new images.

The current generation of tools generate images using a process called diffusion. Before this takes place, the model is trained using a dataset of hundreds of millions of images, scraped from places like google images (along with their text descriptions, the 'alt sets'). The important thing to note is that the new image does not come from the training data: the generator doesn't just stitch existing images together. It comes from the 'latent space' of the deep learning model, which we can understand simply as a multi-dimensional, mathematical space in which the model sorts through images. Within the latent space, the training model 'learns' how to differentiate between images, by drawing out variables such as colour and shape. When a prompt comes through, the model will then match this prompt with an area of the latent space.

This is where diffusion comes in, which means that the image will start with a randomised set of pixels, before being added to with the prompt results from the latent space. The process has been described in artistic terms as "pointillism in reverse", since instead of starting with a complete picture and fractionalising it into thousands of composite parts, the process begins with the

pixels and gradually builds them up into something familiar.

Why now?

The virality of these generators has been evident since the summer. The most prominent images tend to be artistic, hyper-real or humorous – or some combination of the three. Having addressed how these generators work, it is worth understanding why they have all appeared with such gusto.

It is most straightforwardly explained in terms of compute power: they are bigger and better. The first AI breakthroughs of the late 2000's came from both computer's ability to run very large neural networks and the provision of millions of data points from the internet to train them. The leap from these earlier 'analytical' AI breakthroughs – i.e. predicting the ETA of your delivery – to the current 'creative' model embodied by DALL·E 2 turns on size. Compute has become cheaper, and the new techniques such as diffusion models have massively shrunk down the costs required to train AIs. Concurrently, researches have created better algorithms.

A straightforward example is the OpenAI's GPT-3 from 2020, versus its 2019 GPT-2 model. GPT stands for 'Generative Pre-trained Transformer' and it is a language model that uses deep learning to produce human-like text. GPT-2 was 'fed' 40 gigabytes of data (7,000 unpublished books) and had 1.5bn 'parameters'. GPT-3 was given 570 gigabytes – made up of more books and the entirety of Wikipedia, among other things – and boasted 175bn parameters. Its training required far more computing resources, but it hugely outperformed its forerunner, creating credible writing samples – _such as this Guardian article convincing us robots come in peace._ It is now used for writing code: Codex and Copilot, services based on GPT-3 are successfully able to turn programmers' natural language descriptions of what they want into actual code.

GPT-3's ability to simultaneously be a journalist and coder is indicative of the increasing breadth of tasks AI systems can take on, which explains the recent explosion of commercial interest. What until recently appeared as advanced meme generators are now seen as the business tools of the future. _Jake Fawkes_, a machine learning PhD student at Oxford University, explained this was down to the fundamental flexibility of these models. Whilst earlier breakthrough AI systems could solve one specific task, this current generation are known as 'foundation models'.

Sequoia Capital puts this simply in a recent commercial overview of the sector:

'The fields that generative AI addresses—knowledge work and creative work—comprise billions of workers. Generative AI can make these workers at least 10% more efficient and/or creative: they become not only faster and

more efficient, but more capable than before. Therefore, Generative AI has the potential to generate trillions of dollars of economic value'

Whilst image and text generators may have dominated the conversation given their instantly impressive results and open-source applications such as _hugging face_, the plethora of tools which are now ready to disrupt roles including graphic design, code design and video editing is remarkable. One London-based example is _Colossyan AI_, which allows you to create videos with AI actors (image below).

Take the Media & Advertising: the ability to automate agency work and improve ad copy is one use case. There are great opportunities here for multi-modal generation that pairs sell messages with complementary visuals.

So where's the evidence of this? Where are the salivating venture investors? Much hype accompanied the recent _$101M funding raise for open-source platform Stability AI_ – with the New York Time's describing a recent event hosted by the company as a 'coming out party for Generative AI'. Stability AI is the creator of the Stable Diffusion – a generator differentiated by its open-source nature. Stability AI has made public all of the details of its AI model, including its weights. There is a palpable sense that generative AI is in, and web3 and the metaverse are out. SBF's spectacular fall from grace with FTX's recent capitulation only served to underscore this.

Commercial barriers

But there are issues. The underlying tech may be there, but not everyone is on board: worries about the possible uses and effects of these tools provoke fierce debate. These range from the proliferation of child-porn, to the death of the artist, and the projection of ludicrous levels of racist and sexist bias.

The Imagen team in particular are acutely aware of the potential pitfalls of AI imagery, with an explicit statement about "Limitations and Social Impact" on their research page. "There are several ethical challenges facing text-to-image research broadly," the statement reads. "First, downstream applications of text-to-image models are varied and may impact society in complex ways … Second, the data requirements of text-to-image models have led researchers to rely heavily on large, mostly uncurated, web-scraped datasets [that] often reflect social stereotypes, oppressive viewpoints, and derogatory, or otherwise harmful, associations to marginalized identity groups."

The academic community is particularly vocal in their opposition to the widespread use of these generators due to the data they are trained on. A widely cited paper by Abeba Birhane, Vinay Uday Prabhu, Emmanuel

Kahembwe entitled _Multimodal datasets: misogyny, pornography, and malignant stereotypes_ spells out some of these concerns. The introduction points out: '[W]e found that the dataset contains troublesome and explicit images and text pairs of rape, pornography, malign stereotypes, racist and ethnic slurs, and other extremely problematic content. We outline numerous implications, concerns and downstream harms regarding the current state of large scale datasets while raising open questions for various stakeholders including the AI community, regulators, policy makers and data subjects.'

One only has to type in prompts such as 'nurse' on Stable Diffusion, for example, to be greeted exclusively with images of women or 'CEO' to an eery approximation of the average middle-aged white man. One of the few professional prompts that could generate any non-white figures was 'fast food worker'. There are also more bizarre episodes which demonstrate the potential of these tools to create wholly unexpected and unpleasant images, threatening their commercial use: this _twitter thread_ details the accidental and unexplained generation of horrific imagery.

These should be understood as cultural _and_ commercial issues: no one will want to risk using generative AI in their business if the data they are trained on enforces damaging stereotypes. These aren't teething problems, but fundamental issues. Companies such as _Unitary_ are building contextual content moderation solutions by using machine learning models to analyse visuals, text and audio together to interpret content in context, but it has yet to be applied meaningfully to generative AI.

The death of the artist?

Part of the wider cultural resistance to generative AI centres on the threat it poses to creative livelihoods. Who needs artists, writers or illustrators anymore when you've got DALL·E 2 and GPT-3 freely available. This was felt particularly keenly with the publication _of an article in the Atlantic_ earlier this year by contributing writer Charlie Warzel about Alex Jones. To illustrate the article he used an image of Alex Jones he'd generated by Midjourney with the caption _Alex Jones inside an American Office under fluorescent lights. AI art by Midjourney._ The image works almost by accident; its blurry imprecision adding to the general chaos surrounding Jones at the time.

Soon after being put online, he received a deluge of hate from people online accusing him of taking away work from illustrators and artists. This panic felt by the use of these images has now subsided – much of it was the shock factor, although undoubtedly this will continue to encroach on illustrator's livelihoods. But there are more nuanced points here: DALL·E 2 is trained on

images taken from innumerable creatives stretching back millennia. It is therefore not a stretch to argue these generators are plundering the history of human creativity to make a commercial product.

These fears also tap into the wider debate on the existential threat of AI. The more advanced the machines become, the higher the chance of human bondage. Generative AI adds a unique aspect to this argument because it takes us on not at mapping out a route to work, or even beating us at chess – but creating original works of art. Art, often seen in opposition to science, has been co-opted by machines.

This existential fear is paraded most vehemently by 'long-termers': people who adopt an ethical stance which gives priority to improving the long-term future. Long-termism is a central tenet of the feted organisation _Effective Altruism_, whose prophets include Sam Bankman-Fried and the philosopher William MacAskill. These thinkers quantify morality precisely; working out exactly what deserves time and resources by reference to the size of the threat it poses. MacAskill recently highlighted a _survey_ in which machine learning researchers predict there is a 50% chance of "human level machine intelligence" within 40 years, and that we are more likely to die from this than a car crash. There is a direct link between generative AI and long-termism in the form of _OpenAI_, the organisation behind DALL·E 2 whose founders including Elon Musk and Sam Altman are motivated by concerns about the existential risk from artificial general intelligence.

This is all perfectly valid. But the fear long-termism generates distorts the discussion of 'the end of art'. Art is a fundamentally human pursuit and any suggestion these generators are replacing artists is not true. Let us think about impactful art in 2022: it is not pure mimicry of other artists; it is not odd images of Alex Jones in his basement; it is not posters of Soviet Afrofuturism. Impactful art is collectives such as _Forensic Architecture_ – who act as an investigative taskforce, creating videos and aesthetic presentations of state violence and human rights violations, to correct official narratives. Photography didn't finish off painting – in fact it became an indispensable tool for some of its foremost practitioners – Francis Bacon, for example. Good art requires a distance from its source material; a kind of shamanistic transformation. Generative machines cannot do this: they can only reorder and reconfigure that which has happened.

These generators will eventually be embraced as useful ideating tools for artists: allowing them to create quick mood boards or digital sketches. In the way marketers are not replaced by Canva, but empowered by it, AI generation is simply 'templatising' creativity.

Article by MAX LUNN