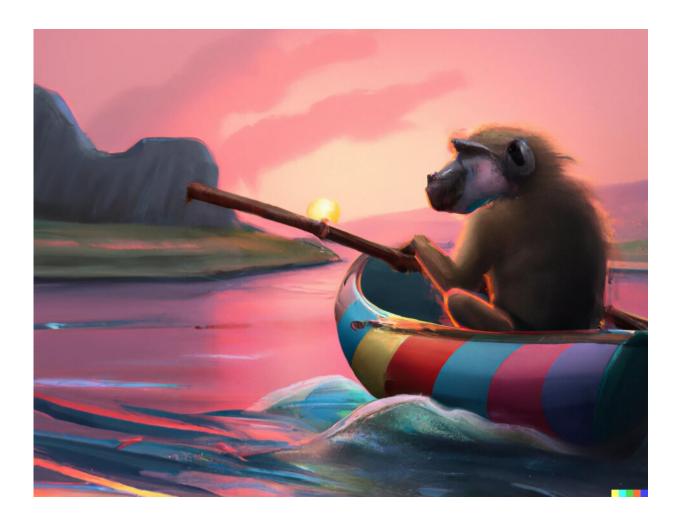# What DALL-E reveals about human creativity

January 19 2023, by Gordy Slack

DALL-E's version of a cute baboon sailing a colorful dinghy at sunset. Credit: DALL-E

The often delightful and arresting images created by the latest generation of text-to-image generators, exemplified by DALL-E 2, Midjourney, and Stable Diffusion, have stirred up lots of buzz in both the arts and the AI worlds. The images, generated from simple text prompts (e.g., a baboon sailing a colorful dinghy), look very much like the products of intelligent human creativity.

To explore just how creative these models really are and what they can teach us about the nature of our own innovative propensities, we asked four authorities on [artificial intelligence](#), the brain, and creativity (and we also asked GPT-3, a language-generating model that's a close cousin to DALL-E) to explain what they think of DALL-E's capabilities and artistic potential.

How DALL-E Works

DALL-E starts by taking billions of bits of text from the internet and translating them into an abstraction, which it stores in a location in "latent," or logical, space. In the universe of describable things, for example, "baboon" will be "located" by strong associations near to other primates, probably not far from "Africa," "savanna," or "zoo." Images, too, are read from the internet and are associated with their captions and transposed into the same logical areas. So, the text and the relevant descriptions of the images, while still distinct, are located by strong associations near to each other. This allows DALL-E to find the kinds of images in the spaces indicated by the user's text prompt. It then generates a set of key features that it has learned this image might include. In our "baboon in a dinghy" example, it would come up with characteristic features for baboon, say the color of its fur, its human-like arms and hands, or the canine shape of its head, as well as characteristic features of a dinghy, say the curved gunwale. Then, DALL-E deploys what's called a diffusion model, which starts with static noise and then sculpts the pixels in a manner informed by the latent representation of

the text description, thus building unique images each time the program is run.

The first diffusion model was invented at Stanford in 2015 by Jascha Sohl-Dickstein, now a research scientist in the Brain group at Google. Seven years ago, when Sohl-Dickstein was a postdoc in the Neural Dynamics and Computation Lab, he and the lab's director, neuroscientist Surya Ganguli, Ph.D., "were exploring ideas in non-equilibrium thermodynamics," says Ganguli. "That work lead to the idea that one could reverse the flow of time in a diffusion process that turns data into noise by training a neural network, which could then turn noise into data," Ganguli says.

Inspired by the brain

Isaac Kauvar, Ph.D., a Wu Tsai Neurosciences Institute Interdisciplinary Postdoctoral Scholar working in the Stanford Autonomous Agents Lab at the intersection of AI, neuroscience, and psychology, points to two analogies between the way DALL-E generates its images and how creative human artists make theirs. The most obvious is that DALL-E is built around a software architecture known as a "neural network" that in concept, if not in detail, mimics the brain's composition of neurons, each one of which has connections to many others. Those connections can be strengthened or weakened during learning, thereby forming meaningful patterns of associations.

What's more, "at a high level, the way that DALL-E builds images from its own latent space is not entirely dissimilar to the way human brains might store and identify concepts and then translate them into outputs," says Kauvar. These abstract concepts help us link, say, the word baboon to an array of different associations and images—colorful bottoms, zoo enclosures, the African savanna.

'Cartoon' version of human creativity?

Neuroscientist, Wu Tsai Neuro affiliate and author David Eagleman, Ph.D., agrees that models like DALL-E do have at least one thing in common with human intelligence: They work by "absorbing a lot of examples and then generating new things based on combining and recombining them," he says. "Creative people also absorb the world, generate remixes, then make whole new versions."

But, when it comes to creativity, says Eagleman, "what these image generators lack is at least as important as what they share with us. That is, they do not have any way to filter what is good, let alone what is profound or beautiful."

Eagleman calls the way image generators learn and produce art "a cartoon version" of the way humans do these things. For one thing, he argues, it's not enough just to make new things. To be fully creative, a person—or a creative machine—would have to be able to filter those new things and select the most resonant and relevant based on human criteria, he says. "DALL-E can't do that. It has novelty down, but not the filtering, the selectivity," he says. "It would have to learn what it is to be a person before it could filter based on human criteria, before it could know whether or why humans would appreciate a particular drawing."

"These AIs are so impressive," Eagleman says, "but they're not doing what the human brain does. Not at all. They use very different techniques to get weirdly similar and often wonderful results. But what's most interesting may be just how something so unlike a human can come up with such impressive results."

Honoring the creative process

Kauvar, who is a visual artist, points to another key difference between the way DALL-E works and the way many people do. "When I'm drawing," he says, "it's an iterative process. I usually don't know where I'm going to end up. I first just get something down and that inspires the next iteration, and that inspires the next one, and so on. DALL-E, on the other hand, determines what to draw and then goes straight to making that thing at once in a few seconds. DALL-E can quickly produce many variations, but it relies on a human to evaluate or modify them."

The importance of honoring process in the human act of artmaking is something that Michele Elam, Ph.D., William Robertson Coe Professor of Humanities and Institute for Human-Centered Artificial Intelligence faculty associate director, also identifies as a key difference between a human's creativity and a machine's. Artists value the creative process, considering it a key part of the act of creation and even essential to the meaning of the artwork itself, she says.

"The idea that something like DALL-E could 'free your creativity' by just making it faster and simpler to get a usable product suggests that artists are burdened by the thought, reflection, experience, care, and time that go into their work," she says. "But for many artists, the meaning of the work is an expression of those efforts, of that process, not incidental to them."

Acting, not thinking

Computational neuroscientist Manish Saggar, Ph.D., Wu Tsai Neuro and HAI affiliate and an assistant professor of psychiatry and behavioral science, has studied human creativity and the brain for more than a decade. In a 2017 paper in *Cerebral Cortex*, he found that one measurable quality of a brain in a highly creative state is the simultaneous deactivation of the right prefrontal portion of the cortex and the increased connectivity between many disparate regions of the

brain. That increased connectivity includes communication between the prefrontal cortex and the cerebellum which, among other things, Saggar says, can be thought of as the brain's graphical processing unit, or GPU. "It's like the CPU and its inhibitions are shutting down, and the GPU is taking over the creative work." Insofar as that is like a withdrawal from strict executive control and a shift to a more distributed, image-based process, it may be broadly analogous to what DALL-E does, Saggar says.

Saggar's team also found that most extraordinarily [creative people](#) have a strong bias toward action. They don't just (or even mainly, at first) think about what they might draw; they simply take pen to paper and start drawing. "Think less, do more" is good creativity-inducing advice, says Saggar.

Likewise, a text-to-image model can be so generative perhaps because it's not trying to force preconceived expectations and apply traditional approaches to a problem; it is simply looking for patterned associations and giving them a try. DALL-E thinks not at all … it only does.

But at some point, argues Eagleman, to complete the creative process, one has to employ what he calls "the human filter" to identify among all those new things the best ones and then, once in a while, to keep working with a favorite new thing until it also becomes a truly great one. As startling and impressive as these powerful text-to-image generators are, they aren't yet close to being artists in that fully creative sense.

When we asked DALL-E's older cousin GPT-3 about the differences between human and AI creativity, it offered the critique that humans have at least one unique selection filter that algorithms don't: "An important way in which humans are still better at generating new ideas is that humans have emotions … . Emotions help to identify which ideas are good and which are bad. They give the motivation to pursue some ideas and not others."

Still a human experience

Like the human brain, DALL-E can generalize from specific ideas or prompts to broader webs of association, allowing it to create convincing [images](#) based on its "experience." And it can also combine concepts in ways that strike us as amusing or creative. We asked it to put the baboon on a dinghy on the ocean, for instance, and to portray it in the style of Georges Seurat; it created a contemplative and lonely-looking baboon adrift in a pointillist sea of colorful dots.

But there are a couple of things central to human creativity that DALL-E still lacks. For one, it has no emotional evaluation of what makes an image important, novel, funny, or meaningful. Perhaps related to this is that DALL-E has no extended artistic process. For human artists, that process is central—trying things, evaluating them, iterating to the next version or the next idea to ultimately discover or zero in on the artist's impetus for making the art in the first place.

For these reasons, full creativity remains—for now at least—in the realm of the human. Perhaps that's why the best products of DALL-E reflect a partnership between the algorithmic image generator and the creativity, selectivity, and insight of a human creator or artist who is wielding it.

Provided by Stanford University

Citation: What DALL-E reveals about human creativity (2023, January 19) retrieved 25 April 2024 from https://techxplore.com/news/2023-01-dall-e-reveals-human-creativity.html