РЕГИОНЫ

Запись 2 (4:22)

WHAT AI IMAGES REVEAL ABOUT OUR WORLD

Although there were plenty of precursors, it wasn't until January 2021 that talk of AI artists became big news, as people began to learn of the image-generating platform Dall-E. Back then, descriptions of the "AI artist" still felt like something out of a children's book: type in a sentence and the computer magically spits out an image!

The technology sounded too advanced to be real, but it had been coming down the pipeline for decades. The first neural network was proposed in 1943, and the technology's development continued in fits and starts throughout the 20th century. As early as 1989, neural networks could decipher typed and handwritten characters, and computer-vision applications expanded rapidly as hardware capacity increased. Soon, optical character recognition allowed us to convert PDFs to editable text, and now we can copy text snippets in photos taken on our phones. Optical character recognition relies on natural language processing, the field concerned with enabling algorithms to output and receive messages in human language rather than a programming language. Natural language processing combines computational linguistics with statistical modelling and algorithms – now usually neural networks – to process and produce "natural" language through methods such as breaking down sentences, tagging parts of speech, assessing words' most frequent positions in a sentence, and highlighting words that do the most prominent signifying (usually nouns and verbs).

By 2015, algorithmic processes were able to form simple sentences or phrases to describe an image. Patterns of pixels identified as, say, "cat" or "cup" were matched with linguistic tags, which were then translated into automated image captions in natural language. Quickly, researchers realised they could flip the order of these operations: what would it look like to input tags – or even natural language – and ask the neural networks to produce images in response? But reversing the image-to-text operation proved less than straightforward, as there's a vast difference between the complexity of a basic phrase and even the simplest image. (While almost any image of a large, centred feline could be described as "a closeup of a cat", there are infinite possible ways to depict the phrase.) One would also have to collect an enormous quantity of visual data to build up an understanding of the near-infinite visual signs that can be described in language.

Some early attempts at image generation dealt with the problems of complexity and dataset size by constraining both the style of an image and its subject matter. The authors of a pivotal 2016 paper – Generative Adversarial Text to Image

Synthesis – began by training their models on limited libraries of images, specifically the Oxford-102 Flowers and Caltech-UCSD Birds datasets.

The bird dataset contains 11,788 photographic images of birds broken down into 200 mostly North American species, annotated with additional attributes such as “Bill Shape”, “Belly Pattern”, and “Underparts Color”. The dataset’s images were downloaded from Flickr and then categorised and annotated by human workers hired on Amazon’s Mechanical Turk, a crowdsourcing platform often referred to as “artificial artificial intelligence”. While one might assume today’s text-to-image tools have been automated all the way down, their architecture and maintenance rely on enormous quantities of human labour, whether the repetitive “clickwork” performed predominantly in the global south by workers paid pennies per “task”, or the voluntary, quotidian labour you’ve provided each time you’ve filled out a Captcha.