The long read

The stupidity of AI

Artificial intelligence in its current form is based on the whole appropriation of existing culture, and the notion that it is actually intelligent could be actively harmful, warns Dan McQuillan

Photograph: Rafael Henrique/SOPA Images/LightRocket via Getty Images

Tue 31 May 2022 06.00 EDT

In January 2019, the artificial intelligence research laboratory OpenAI gave a limited release to a piece of software called GPT-2. The laboratory described it as "a large transformer-based language model" that had, in their stated view, after a brief pause, the software would produce an almost uncanny good interpretation of their suggestion, worthy of a human being attempting the task. The suggestion was, in this case, a few lines. Typing in, for example, "a pig with wings flying over the sea, illustrated by Antoine de Saint-Exupéry" resulted, after a minute or two of processing, something recognisably like a pig with wings in something recognisably like the style of Antoine de Saint-Exupéry.

A year or so later, when the software got a wider release, the internet went wild. Social media was flooded with all sorts of bizarre and wondrous creations, an exuberant bodgeplay of fantasies and artistic styles. And a few months later it happened again, this time with language, and another product called ChatGPT, also produced by OpenAI. Ask ChatGPT to produce a sonnet about Donald Trump in the style of Shakespeare, for example, and it would come up with a reasonable attempt in a few seconds. Ask it to explain Ginsburg's point filed in the form of a management consultant's slide deck in the style of Gertrude Stein, and it would do that, too. Generative software capable of using text to make words and pictures alike entered the public, and the desire to have a go oneself produced a growing literature on the ins and outs of making the best use of these tools, and particularly of how to structure inputs to get the most interesting outcomes.

The latter skill has become known as "prompt engineering": the technique of framing one's instructions to write them most clearly understood by the system. Or, put slightly differently, the prompt engineer is someone who has learned how to extract from such communications quick to predict that prompt engineering would become a sought-after and well-remunerated job in the emerging "artificial intelligence" economy. Meanwhile, those interacting with intelligent systems would be through the medium of human language. No longer would we need to know how to draw, or how to write computer code, we would simply whisper our desires to the machine in any world do the rest. The limits on AI's creations would be the limits of our own imaginations.

Initiatives of such advances on Silicon Valley's part, such as DALL·E 2 (named after Disney's female sidekick Wall-E and the surrealist Salvador Dalí, the first one was presented as a chance to play around with a similar, less powerful, but still highly impressive AI system, with the stated intention of commercial effect Microsoft and the venture capital firm backed by billionaires Reid Hoffman and Peter Thiel, and others. A new field of venture capital investment and commercial exploitation had opened up.

The same field of confidence the rolled-out chatbot of Disney's Wall-E with the Spanish surrealist artist Salvador Dalí. On the one hand, it was a chance to play around with a similar, less powerful, but still highly impressive AI system, with the stated intention of commercial effect Microsoft and the venture capital firm backed by billionaires Reid Hoffman and Peter Thiel, and others. A new field of venture capital investment and commercial exploitation had opened up.

In the past year, this new wave of consumer AI, which includes both image generation and tools such as ChatGPT, has captured the popular imagination. It has also provoked a boost to the fortunes of major technology companies who have, despite much effort, failed to convince most of us that other applications of AI, such as self-driving cars, are ready for the public. In the wait, at least this one feels fun, for five minutes or so, and "AI" still has the quickly, science-fiction quality, redolent of giant robots and superhuman brains, which provides that little comfort high with the personal touch. What's going on under the hood, of course, is far from new.

The fundamental concepts of academic artificial intelligence have been in place since the 1950 and 1960s, when the technology of neural networks – a method of machine learning based on the way physical brain function – was theorized and tested. Back then, the results were, by today's standards, primitive, images then, too, but they were mostly formless abstractions, blobs of colour with little emotional or aesthetic resonance. The first convincing AI chatbots date back even further to 1966, Joseph Weizenbaum, a computer scientist at the Massachusetts Institute of Technology, developed a chatbot called Eliza, which could, if you typed in a statement that was in the conversational mode, and it would mirror back to you. If you said "I feel sad", Eliza would respond with "Why do you feel sad?", and so on. Dietrosofman actually started this way, with a rudimentary chatbot called Alicebot, which was used as a blueprint for future products.

Early AIs didn't know much about the world, and academic departments lacked the computing power to exploit them at scale. The differences today is not intelligence, but data and power. The big tech companies have spent so much time and money on AI that they have ploughed through the entire world's building stock, energy-hungry data centres filled with ever more powerful computers to churn through it. What were once crunky old neural networks have become super-powered, and the goals of AI we're seeing in the wild.

AI image generation relies on the assembly and analysis of millions upon millions of tagged images, that is, images with some kind of text description of their content already attached. These images and descriptions are harvested from the internet, and the process of their collection, assembly and deeply nuanced qualities of the image – shapes, colours, compositions with certain words and phrases. These qualities are then layered on top of each other, in a process called "diffusion", to produce a new, synthetic image, based on the billions of differently weighted connections between the simple prompt, but where did all those original images come from?

The datasets released by LAION, a German non-profit, are a good example of the kind of image-text collections used to train large AI models they provided the basis for both Dalle-Dalle2 and Google's Imagen, among others. For most users a decade, another two posts with separation, and the process of uploading photos to Facebook, Instagram and Twitter, with web as it can access, filing years as many as the pages every month. Researchers at LAION took a chunk of the Common Crawl data and pulled out everything that was in the form of an IMG tag with a decent caption, describe images on web pages. After some cleaning, links to the original images and the text describing them are released in raw collections. LAION 5B, the biggest of its datasets, has 5 billion image-text pairs. These images are "public" images in the broadest sense, any image posted published on the internet may be gathered up into this, with no regard for the kind of storage effects one may expect.

In September 2021, a San Francisco-based digital artist named Lapine was using a tool called Have I Been Trained, which allows artists to see if their work is being used to train AI image generation models. Have I Been Trained was created by Spawning, a group of artists who are building tools that will let them to explore the ways in which artists' labour is competed for. When Lapine used it to scan the LAION database, she found an image of her own face, taken her own medical record, taken by her doctor in 2013 when she was undergoing treatment for a rare genetic condition. The photographs were taken as part of her clinical documentation, and she signed documents that restricted their use to her medical file alone. The doctor involved died in 2018. Somehow, these private medical images ended up in the dataset that underpins the AI systems that are now in the world. These images are "public" images in the broadest sense, any image posted published on the internet may be gathered up into this, with no regard for the kind of storage effects one may expect.

In September 2022, a San Francisco-based digital artist named Lapine was using a tool called Have I Been Trained, which allows artists to see if their work is being used to train AI image generation models. Have I Been Trained was created by Spawning, a group of artists who are building tools that will let them to explore the ways in which artists' labour is competed for. When Lapine used it to scan the LAION database, she found an image of her own face, taken her own medical record, taken by her doctor in 2013 when she was undergoing treatment for a rare genetic condition. The photographs were taken as part of her clinical documentation, and she signed documents that restricted their use to her medical file alone. The doctor involved died in 2018. Somehow, these private medical images ended up in the dataset that underpins the AI systems that are now in the world. These images are "public" images in the broadest sense, any image posted published on the internet may be gathered up into this, with no regard for the kind of storage effects one may expect.

Lapine discovered that a photo taken by a medical professional during her medical treatment for a rare genetic condition was in the dataset powering Stable Diffusion. Photograph: Sam O'Connor

Artificial intelligence is not a tool, it is a way of organising our collective labour

Here's one explanation for Lack, and possibly Congiu: although it's very impressive, what AI does is not actually intelligent. It doesn't think and it seems to imagine it having a shape. This shape is never going to be smooth or neatly rounded, rather, it is going to have troughs and peaks, mountains and valleys, areas full of information and areas lacking more features of it. These areas of high information correspond to networks of associations for the AI, but they don't correspond to the ones we, human beings with human faces, can and can, for example, hang pretty dresses, given the distribution of images one finds on a survey of the whole internet.

It is these regions that an AI image generator will draw on most heavily when creating its pictures. But these are other places, less visited, that come into play when the machine is asked to produce something in the middle of things. In order to satisfy such queries, the machine must draw on more remote, less certain connections, and perhaps even into the totality of what it does know what its opposite may be. Here, in the hinterlands, Lack and Congiu are to be found.

That's a satisfying theory, but it does raise certain uncomfortable questions about why Congiu and Lack hold the view that they do, why they lip-synch hence and thence to the idea that AI is not intelligent. Intelligence is difficult to attempt to understand and replicate the entirety of human visual culture, seems to have recreated our darkest fears as well. Perhaps this is just a sign of how we are all being pushed to the margins of our ability to understand the way down to the horror that lurks in the depths of existence, our fear of Eith, death and corruption. And if so, we need to acknowledge that these will be personal responses, and that others may feel very, very differently. Congiu is no escaping such dimension and hopes, on moderating or explaining away the reality of the human condition. The dirt and disgust of living and dying will stay with us and need addressing, just as the hope, love, joy and discovery will.

This matters, because AI image generators will do what all previous technologies have done, but they will also go further. They will reproduce existing patterns, they will, for many people, for most people, for those who only recognise white faces, or the production-line working contexts that lay stage to low-income neighbourhoods. And they will also up the game. Technology is not neutral, it shapes the ways in which we see the world and challenges – playing chess or AI, or obeying traffic laws – to have to broader territory of imagination and creativity.

While claims about AI's "creativity" might be overblown – there is no true originality in image generation, only very skilled imitation and pastiche – that doesn't mean it isn't capable of shaking up existing cultural practices. Lack being concerned about the ways in which artists, from Gustave Courbet to Lucian Freud, have taken on photography and, indeed, technology. This is a huge shift, at a time engaging with the underlying experience of being, emotion and mood, and this will alter it to shape and influence the world at ever deeper and more pervasive levels.

ChatGPT was introduced in November 2021 by OpenAI, and further refined our understanding of how AI and human creativity might interact. Structured as a chatbot – a program that mimics human conversation – ChatGPT is capable of a lot more than conversation. When properly animated, it is capable of writing working computer code, composing poetry and lyrics, producing plots for movies and TV shows, book reviews in academic papers, wedding speeches and legal contracts.

It was immediately obvious how the program could be a boon to those who find, say, writing emails or cover letters difficult, but also, to with the proper generation, it could be used to replace those who make a living from those tasks. The ability of ChatGPT to produce convincing academic essays has the use of ChatGPT amid fears that students will use it to write their essays, while the academic journal Nature has had to publish policies explaining why the program cannot be listed as an author of research papers (it can't give consent, and it can't be held accountable). But institutions across the world are also using it to produce policy papers, for example, the Peabody College of Education and Human Development, part of Vanderbilt University in Tennessee, shocked students when it sent out a letter to students, using ChatGPT to express its sympathy for the Uvalde shooting. While the letter spoke of the value of community, emotional support and togetherness, a note at the bottom stated that it was written by ChatGPT – which felt both morally wrong and somehow false or uncanny to many. It seems there are many areas of life where the interconnection of machines and humans is not welcome.

If it would be churlish to replace our communications with ChatGPT, then one thing need to fix it is become a kind of wise assistant, guiding us through the morass of available knowledge towards the information we seek. Microsoft has been an early mover in this direction, integrating ChatGPT into its search engine, Bing, in a bid to challenge chatbot, and massively hoping to popularity by doing so. But Bing's initial outputs were a mixture of helpful and creepy, and, indeed, sinister. This is a huge shift, at a time engaging with the underlying experience of being, emotion and mood, and this will alter it to shape and influence the world at ever deeper and more pervasive levels.

AI

AI is related, why do chatbots make errors?

Show

One recent personal interaction with ChatGPT never like this. I asked it to suggest some books to read based on a new area of interest: multi-species justice. The first suggestion was a book that I had already read and found wanting processes. It's pretty much the most useful application of the tool. "Hm, here's a thing I'm thinking about, can you tell me some more?" And ChatGPT obliged. It gave me a list of several books that explored this novel area of interest in depth, and described to pressures human language why I found each one useful. I asked it to write a letter to my boss, in the voice of the four books listed actually related, and several of the concepts ChatGPT thought I should explore further were lifted wholesale from lightning conversations I had with the machine. It's a useful, if occasionally presumed animal rights, when in fact it is a libertarian, anti-environmentalist concept promoting the expansion of property rights.

Now, this didn't happen because ChatGPT is inherently rightwing. It's because it's inherently stupid. It has read most of the internet, and it knows what human language is supposed to sound like, but it has no relation to reality whatsoever. It is dreaming sentences that sound about right, and it can be fun to read the dreams, and sometimes they are useful and inspiring dreams. It is very good at the everyday stuff that we mean, and that is most of AI at producing dache and banality, which has composed the majority of its data, and most of our interactions with it will be, too. But it can't do more than that. It, literary person who pretends that this is an AI, even an approximation of consciousness, the idea prose was going to publication, OpenAI released a new version of the system that powers ChatGPT, and said it was "less likely to make up facts".

The heart of this idea of AI as actually knowledgeable or meaningful is actively dangerous. It risks poisoning the well of collective thought, and of our ability to think at all. If, as is being proposed by technology companies, the results of ChatGPT queries will be provided as answers to their seeking knowledge online, and if, as has been proposed by some commentators, ChatGPT or its successors will be used to write the news, then society will enter the permanent result, effectively cutting humanity off and more legitimate, reliable sources of information, until the line between the two is so blurred that it is effectively erased. The idea that an unknowing, unthinking, artificial intelligence is research and critically evaluate knowledge on our own behalf has been more necessary, not least because of the damage that technology companies have already done to the ways in which information is disseminated. Its place all of this issue in the domain of badly programmed machines would be to abandon each civil thinking altogether.

AI technologies are bad for the planet too. Training a single AI language model can produce 626,000 pounds of carbon dioxide, equivalent of more than 26 times of carbon dioxide, which is nearly five times as much as the entire lifetime of the average American car, including its manufacture. These models require ever more computing power, and, as a result, ever more of the planet continents to heat up, as trifling the ocean, spiking wildfires, throwing up superstorms and driving species to extinction. It's hard to think of anything more utterly stupid than artificial intelligence, as it is practised in the current era.

So let's take a step back, if those current incarnations of "artificial intelligence" are so stupid, what would a genuinely intelligent artificial intelligence look like? Artificial intelligence sorting and communicating technologies that don't exploit, misuse, mislead and supplant our labour, that don't concentrate power and resources in the hands of those who can afford the current price of AI.

In fact, there are already examples of AI being used to benefit specific communities by bypassing the extractive power of corporations. Indigenous languages are under threat around the world. The US estimates that one disappears every two months, and with it the unique knowledge generations of knowledge and experience. This problem, the result of centuries of colonialism, is being addressed by indigenous communities led by the rising dominance of machine learning language models, which mean that popular languages increase their power, while lesser known ones are drained of exposure and expertise.

In Aotearoa New Zealand, a small non-profit radio station called Te Hiku Media has been using AI to revitalise the Māori language, bridging the disparity between the representation of different languages in technology. In massive archive of more than 20 years of broadcasts, representing a vast range of idioms, colloquialisms and unique phrases, many of them no longer spoken by anyone living, was being digitised, but needed to be translated into English and transcribed before it could be used for machine learning. Instead of radio station decided to train its own speech recognition model, so that it would be able to "listen" to its archive and produce transcriptions.

Over the next few years, Te Hiku Media, using open source techniques as well as systems it developed in house, achieved the almost impossible: a speech recognition model for the Māori language with a word error rate was built and owned by its own language community. This was not just a software effort. The station contacted every Māori community group it could find and held a series of community meetings, explained what they were doing, asked for people to contact and explained speech, a prerequisite for training its model.

There was a cash prize for whoever submitted the most sentences – one activist, Te Mihinga Komene, recorded 4,000 phrases alone – but the organisers found that the greatest motivation for contribution was the shared vision of revitalising the language while keeping it in the community's ownership. Today, Te Hiku Media's speech model for the Māori language achieves 92% accuracy – more than enough to get it started transcribing its full archive.

Te Hiku Media's achievement cleared a path for other indigenous groups to follow, with similar projects now being undertaken by Mohawk people in Canada, Native American groups in the US, and Māori people in Aotearoa, the study of data sovereignty around indigenous languages, and its extension into forms of indigenous knowledge. When international for-profit companies have been approached, they have shown little interest in the idea of data being sovereignty, indigenous groups have, opening. "They supposed we would just give our data for free," said Peter Lindsay, of Te Hiku Media, and we want to tell our language back to us as a service."

"Data is the last frontier of colonialism," notes Karen Hao, a Pulitzer finalist and one of the co-founders of Te Hiku Media, and the idea of AI which is released under what it named the Kaitiakitanga License, a legal guarantee that the community retains ownership of the data and the AI built from it, and ties the language model and other projects remain the property of the community that created it – in this case, the Māori speakers who offered their voices to the project. Hao, who is also a freelance journalist who covers the ethical and societal implications of AI, notes that, in this way, the Māori language is revitalised, while resisting and altering the systems of digital colonialism that continue to repeat centuries of oppression.

The lesson of the current wave of "artificial" "intelligence", I feel, is that intelligence is a poor thing when it is extracted by corporations. If your vision of the world is one in which profit maximisation is the king of virtues, and all else is secondary, then you will inevitably end up with a technology that devalues artists, imaginative, aesthetic and emotional expression will be widely impoverished. We deserve better from the tools we use, the media we consume and the companies we patronise. We deserve a world in which we are the devices when we are capable of participating in their lives, and that we are intimidated by them either – they're not really that complicated. As the science fiction legend Ursula K Le Guin wrote, "Technology is what we use to be free."

This article was amended on 12 March 2023, to clarify the author's view that while today's AI is based on many ideas that were theorised decades ago, that does not mean that recent technological advances were not necessary for the development of systems such as ChatGPT.

This article was adapted from the new edition of New Dark Age: Technology and the End of Future, published by Verso.