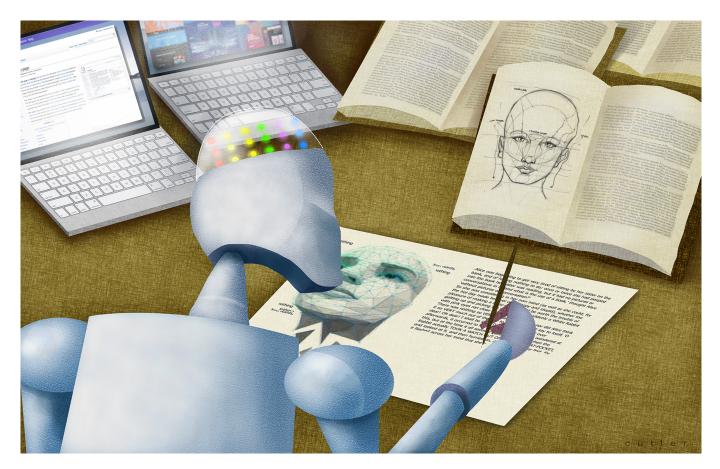# The muse in the machine

As generative AI gets more inventive, what are the implications for human creativity?

David Adam, Science Writer

Fast-improving AI is an impressive tool that could offer new ways to create, but is also a flawed newcomer that could mislead users and even denigrate the creative process. Image credit: Dave Cutler (artist).

Artists have long looked to muses for inspiration. Model and actress Pattie Boyd inspired songs from both George Harrison and Eric Clapton. Oscar Wilde's love for Lord Alfred Douglas encouraged Wilde to pen his famous plays. Even William Shakespeare opened *Henry V* with an earnest cry for creative help: "O for a Muse of fire, that would ascend the brightest heaven of invention."

Enter machines. Can a computer act as a muse? Can an algorithm create art?

To find out, artists and writers are among those experimenting with machine-learning computer models, which are trained on centuries of human works and can produce their own works on demand. From the high-profile language bot ChatGPT to visual-art generators like DALL-E, results suggest that artificial intelligence (AI) can now mimic human creativity at the touch of a button. Or can it?

As researchers, artists, and others assess the fast-improving AI technology's capabilities and shortcomings, they're seeing an impressive tool that could offer new ways to create—but also a flawed newcomer that could mislead users and even denigrate the creative process. While some experts point to apparent examples of computerized creativity, others argue that AI technology will never match the human brain. "AI creativity is telling us more about our own creativity than anything else," says Marcus du Sautoy, a mathematician at the University of Oxford (Oxford, UK) and author of the 2019 book *The Creativity Code: How AI Is Learning to Write, Paint and Think*. "It is a new telescope on the huge creative output that we have produced to date."

Creativity is notoriously difficult to define. Is it enough for a machine to rearrange words if it does not appreciate them? Does modeling, mimicking, and combining existing artistic styles count as doing something new? When does a computer's

note-by-note demonstration of its mastery of the mathematical roots of music tip from predictable to enjoyable?

Those are philosophical as much as scientific questions. But, prompted by the rapid progress and public interest in what algorithms can achieve, mathematicians, psychologists, and AI experts are working to answer them.

Flavors of Creativity

Assessing creativity itself, whether human or machine, takes a little lateral thinking. Like many in the field, du Sautoy borrows a seminal concept first introduced by computer scientist Margaret Boden in 1998, which breaks down creativity into three types (1). The first, which Boden calls combinatorial creativity, involves the novel combination of familiar ideas. Generations of poets and writers have used this to find fame with a neat image or analogy. Think William Wordsworth and "I wandered lonely as a cloud."

The second, exploratory creativity, takes what already exists and pushes the boundaries to extend the limits of what was done or seen before. French painter Claude Monet exploited new pigments to visualize the way light fell on water lilies and, in doing so, helped launch Expressionism. Almost all human creativity is exploratory, Boden says.

Her third branch of creativity is rarer and more mysterious. Called transformational creativity, it breaks rules, changes the game, and demands to be assessed on its own terms. Who says eyes must be painted on either side of the nose? Not Pablo Picasso. Or picture the impact of David Bowie as Ziggy Stardust performing *Starman* for the first time on television in 1972.

du Sautoy argues that AI has achieved all three creativity types. "Midjourney and DALL-E, I think, could be regarded as interesting examples of combinational creativity—the power of AI guided by humans to mix language and visuals to create something surprising," he says. Both programs generate images from textual and often abstract descriptions. Want to design a lawnmower? Here are a thousand pictures of possibilities, one of which is shaped like a dinosaur and another that is made of fruit.

AIs are also well-placed for exploratory creativity, he says, because their training data often hide untapped potential. A music generator called The Continuator does this for jazz. It analyzes notes played by an improvising musician in real time and then continues to play in the same style, exploring new possibilities within that existing framework.

For a case of AI and transformational creativity, du Sautoy points to a much-discussed round of the Chinese game Go, played in 2016 between Lee Sedol, an 18-time world champion, and AlphaGo, an algorithm developed by the company DeepMind. Go requires two players to alternately place black or white stones on a 19 × 19 grid, each trying to surround—and so capture—the stones of their opponent. For centuries, Go masters tended to place early stones on the board's outer four lines. That's a way to gain short-term territorial control while anticipating play shifting toward the center in later moves. But in move 37 of that game, AlphaGo broke with this orthodoxy and placed its stone on the fifth line in.

It might not sound like much, but commentators and Sedol were staggered. Even AlphaGo knew it was doing something extraordinary, calculating the odds of a human player making that move as 10,000 to 1 against. It proved a masterstroke and, some 50 moves later, tipped the balance and sealed the win

for the machine. "Transformational creativity is the tough one, where something new breaks the old system," du Sautoy says. "I would say that Move 37 had that quality because it challenged the previous system of playing, with a radical new move."

Under the Hood

How do the machines do it? Even the most creative algorithms can work only with the material on which they are trained. But these machines, known as generative AI, come in several types that apply what they learn from training data in subtly different ways.

Language bots like ChatGPT typically employ a type of neural network called a transformer, which finds and learns statistical patterns in the order of words on millions of pages of online text. To create an essay, poem, or slide presentation, it computes what the next word should be, based on all the words that have come before (using all those millions of texts on which it's been trained).

The model also has something called a self-attention mechanism, which allows it to pick out the most important features of a user's request. For example, if one asks a language model to describe how "a car is driving down the street when it gets hit by a truck," then self-attention helps the algorithm identify that the word "it" in that request refers to the car, and not the street. That's something the human mind assumes because a street being hit by a truck doesn't make sense—but the AI doesn't know that.

Visual art AIs such as DALL-E tend to use different technology. Called latent diffusion models, these systems compress and manipulate data from existing images to find mathematical ways to generate other images from random noise.

Among many other types of creative AIs, du Sautoy says some of the most successful are called Generative Adversarial Networks. These combine a generator model, which produces works based on a training dataset, with a discriminator model that must try to distinguish "fake" outputs (those made by the generator) from "real" (original examples similar to those in the training data).

The two compete. While the generator offers fake samples that resemble real ones, the discriminator tries to spot the AI-generated outputs. Over time, the generator learns how to make its own output closer to the originals. "The feedback loop in the algorithm means that the algorithm is growing and learning as it plays and creates," du Sautoy says.

But although AI can do things we don't inspect or instruct, in all cases, computer creativity can only follow human creativity, du Sautoy stresses. "AI needs our data to get going," he says, "so it could never really get started without our creative output."

Anxious Artists

As creative computers push into what has previously been a very human domain, their new-found abilities—and how far they might be able to go in the future—are provoking concern and controversy in some quarters. Earlier this year, ethicists even warned that generative AIs risked "the collapse of the creative process" because they devalue art (2).

The debate has real-world implications. The UK Supreme Court is currently considering whether creative AIs should be

granted intellectual property rights to their inventions. Physicist and entrepreneur Stephen Thaler wants to name his machine DABUS as an inventor on patent applications for an emergency beacon and a food container based on mathematical fractals. As the owner of the AI, Thaler argues that he would also own its patents by default. South Africa's patent office agreed, issuing DABUS a patent in 2021, noting that it was "autonomously generated by an artificial intelligence." Other patent-granting bodies, including that of the United States, refused because they require a human inventor.

> **"It is a co-writer with a tendency to go off the rails, but sometimes it was fascinating to let it go off on a tangent a little, collect its ramblings, and piece together a part of a story out of them."**
>
> **—Yudhanjaya Wijeratne**

Practitioners in science and education, meanwhile, worry that researchers and students could deliberately conceal the creative role of an AI. Shortly after the ChatGPT version of the model (since improved with a newer version called GPT-4) was released in November 2022, researchers at Northwestern University in Evanston, Illinois, showed that it could create convincing text for fake scientific abstracts, a third of which were plausible enough to fool human reviewers asked to identify them (3).

One reason that the fake abstracts were convincing was that the AI knew how large an invented patient cohort should be, says Catherine Gao, a physician-scientist who led the work. For an invented study on the common condition of hypertension, ChatGPT included tens of thousands of patients in the cohort, whereas a study on monkeypox (a much rarer disease) had a much smaller number of participants.

Many scientific journals subsequently published warnings to potential authors about using Large Language Models (LLMs) to help write submissions. Although PNAS and *Nature*, for example, now require authors to declare in research paper acknowledgments or "materials and methods" sections any help from AI language models, *Science* went further and banned any AI-generated content, including figures and graphs. The journal demands original work, said *Science* Editor-in-Chief Holden Thorp. "The word 'original' is enough to signal that text written by ChatGPT is not acceptable: It is, after all, plagiarized from ChatGPT," Thorp wrote in a January 2023 editorial (4).

Matthew Cobb, a zoologist at the University of Manchester in the UK, investigated ChatGPT's abilities after growing concerned that students could submit the AI's output in online exams that ask for text-based answers. He asked the AI to write answers on topics including the conflict between science and religion and the parenting behavior of birds. His worry was justified—up to a point. ChatGPT often produced answers that deserved a passing grade, he says. Still, rather than appearing creative, Cobb found that the language had the feel of generic boilerplate text.

Artificial Amateurs

Can AI help write quality creative fiction with the right prompts? To find out, experts at Google last year gave 13 professional writers the opportunity to work with its LLM, called LaMDA. The writers, who included Ken Liu, a multi-award-winning fantasy writer, and Robin Sloan, author of the 2012 bestselling novel *Mr. Penumbra's 24-Hour Bookstore*, were given access to the AI for 9 weeks and asked to use it to craft a story (5). (The resulting stories can be read here: https://wordcraft-writers-workshop.appspot.com/.)

"One of our goals was to help with the creative process," says Daphne Ippolito, a senior research scientist at Google Brain, an AI division of the company. "Not to replace writers, but there's parts of the writing process that are laborious, are boring or hard, like if you have writer's block. And we really wanted to try to address some of these pain points."

The study broke down creative writing into separate tasks, from idea generation and writing sentences to looking up facts and suggesting words and items according to a specific theme, such as rabbit breeds and their magical qualities. "It was almost like having a constant brainstorming partner always there to bounce ideas off," says Wole Talabi, a Nigerian author who took part in the study. "I would put in one or two sentences and then ask it to tell me what happens next. Even if I hated everything it suggested, it kept me thinking through the different alternatives. So I never got stuck."

Yudhanjaya Wijeratne, an author in Sri Lanka, also found the AI to be a useful prompt. "It is a co-writer with a tendency to go off the rails, but sometimes it was fascinating to let it go off on a tangent a little, collect its ramblings, and piece together a part of a story out of them." He sees some pretty big potential. "I strongly suspect, that with a little bit of prompt engineering, we can actually co-write an entire novel this way."

Despite these positives, Ippolito says that most of the writers were disappointed with the creative aspects of the algorithm. "They expected it to be better at generating interesting stylistic things," she says. The Google algorithm, like other language models, rarely surprised or produced something unexpected. "They don't really generate weird text. If they generate weird texts, it's probably because they made a mistake," she says. "And the sort of weird things that a human writer writes is what makes their writing different." Without that human quality, fictional stories written by AIs alone tend to stand out for their low quality. Already, science fiction and literary magazines have complained about receiving hundreds of hopeless algorithm-penned tales from would-be contributors.

Another weakness of the algorithm, Ippolito says, was that it was just too *nice*. That goes for other language models as well, especially those that have been publicly released. "They bias the models to be agreeable and to agree with whatever the human says," Ippolito says. "But if I ask, 'is my story good?', I don't want the answer to always be 'yes.' If my story is bad, I want the model to answer that the story is bad and explain why it's bad."

Other forms of bias in language models affect creativity as well. Early versions were trained on the full range of (often unpleasant) content available online and could easily be coaxed into making anti-Semitic or racist comments. So most developers now actively train models to avoid a range of topics. That's good for their reputation, but less useful for a writer who might want to engage with the darker side of human nature. "The software seemed very reluctant to generate people doing mean things," one of the authors told Google.

These sensitivities mean that LLMs have whole categories of human experience that are off-limits, says Katherine Elkins, who works on AI and creativity at Kenyon College in Gambier, Ohio. "Drugs, sex, murder, violence—all the great stuff of novelists is filtered out," she says. "So, we're really not seeing what it's capable of." For Elkins, there is no doubt that AI can be creative. "I think the harder question," she says, "is when we come to art."

Ghost in the Machine?

Whether or not to judge creative works produced by a machine as art goes beyond assessment of the finished product, Elkins says. It also raises the issue of intentionality. "When I look at my students training to be artists, I think that they feel that the art that they're making is translating their lived experience—that there is an intention to make art behind it," she says. "And obviously DALL-E doesn't have that kind of lived experience."

Or does it? "The tricky thing with all of this, right, is that it's been trained on all of our art that has come out of that experience," Elkins adds.

Visual arts AIs have proven especially controversial recently, as artists discover that their works were used to train algorithms without their knowledge or consent. Online images often come with useful descriptions and captions that help the AI learn to associate the words and pictures—and then generate new images from text prompts. Some artists have fought back against the machines, launching a copyright lawsuit over the use of images and the ability to reproduce unique styles. "Humans are going to want to defend their territory," Elkins says. But the stable diffusion mechanism at the heart of the visual arts AIs could make proving plagiarism difficult, she adds.

Think of how a drop of food coloring or ink dropped into a glass of water spreads and diffuses into random patterns. "Well, here we're starting with the glass of water, with the ink already diffused, and it's like going in reverse," she says. "So, there are no originals; there's no plagiarism."

Some AI pictures have won prizes and sold for big money. In one now-infamous case, a Colorado artist submitted a Midjourney-produced image to an art contest and won. But because of the way these works are generated, many AI experts are reluctant to call the output of such models "art." "I'm quite reserved about calling AI creative or at least comparing it to what artists do, because I know what these models look like from inside," says Imke Grabe, a machine-learning researcher at the IT University of Copenhagen (Copenhagen, Denmark). "They lack an understanding of how the world works. And I think that's a huge part of working as an artist."

For du Sautoy, this is where intentionality is key—and currently missing from machines. "I think that intention in AI creativity will happen, but I believe that will be a signifier of an emerging AI consciousness," he says. "Once an AI has an inner world, it will be compelled to share this with others, and that will lead to the drive to demonstrate that something is going on inside the AI."

Derivative Designs

A common argument against AIs being creative is that they draw heavily on the data on which they were trained. But, as Ippolito at Google Brain points out, writers, artists, and musicians have always done this. "If you look at the famous classical composers, Tchaikovsky steals from the composers who came before him; Bach steals from the composers who came before him," she says. Arguably, creative works of all types are derivative, at least to some extent.

Indeed, even a writer as creative as Shakespeare doused his muse of fire to routinely steal plotlines and scenes from other writers. "So, are we holding models to a higher standard for creativity, and for borrowing from the past, than humans?" Ippolito says. Perhaps these generative AI algorithms aren't squashing human creativity—just pointing out its inherent limits.

These arguments have been brought into sharper focus recently by the release of GPT-4 to ChatGPT subscribers in March, which, according to those who have seen it, offers a more sophisticated writer than its freely available predecessor. "It does seem like it has a longer attention, which means it can keep coherence and write longer," says Annette Vee, an English professor at the University of Pittsburgh in Pennsylvania, who studies the intersection of writing and computation. "And it has a better sense of humor, which means that it processes context and cultural things a little bit better."

The updated chatbot also comes with an interesting new feature: It can analyze and describe images, including why they are funny. "The fact that it can translate the visual input into text along with all of these cultural things about humor is actually pretty impressive," she says. OpenAI hasn't revealed many details on the improved version, but Vee says the new algorithm is likely merging image models with text models.

Such impressive exploits will continue to provoke both amazement and consternation from writers, artists, and researchers of all stripes, as they reconsider what it means to generate a creative work. "Creativity is a moving target, where people kind of very quickly accept, okay, computers can do this, meaning this is not an example of creativity," says Michal Kosinski, a computational psychologist at Stanford University in California. "We shouldn't be judging everything against human standards."

1. M. A. Boden, Creativity and artificial intelligence. *Artificial Intelligence* **103**, 347–356 (1998).
2. N. Eisikovits, A. Stubbs, ChatGPT, DALL-E 2 and the collapse of the creative process, *The Conversation* (2023). https://theconversation.com/chatgpt-dall-e-2-and-the-collapse-of-the-creative-process-196461. (Accessed 16 February 2023).
3. C. A. Gao *et al.*, Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. bioRxiv [Preprint] (2022). https://www.biorxiv.org/content/10.1101/2022.12.23.521610v1. (Accessed 16 February 2023).
4. H. H. Thorp, ChatGPT is fun, but not an author. *Science* **379**, 313 (2023).
5. D. Ippolito, A. Yuan, A. Coenen, S. Burnam, Creative writing with an AI-powered writing assistant: Perspectives from professional writers. arXiv [Preprint] (2022). https://arxiv.org/abs/2211.05030. (Accessed 17 February 2023).