Generative AI: the new era of machine learning

ChatGPT has burst into the public consciousness in a way seldom seen outside the sci-fi realm. But systems like this, which produce content to order, threaten not just jobs but a surge of misinformation.

By Richard Waters

None of these pictures was created by a human being

FT montage/Images generated by Midjourney and Open.ai

Just over 10 years ago, three artificial intelligence researchers achieved a breakthrough that changed the field forever.

The "AlexNet" system, trained on 1.2mn images taken from around the web, recognised objects as different as a container ship and a leopard with far greater accuracy than computers had managed before.

That feat helped developers Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton win an arcane annual competition called ImageNet. It also illustrated the potential of machine learning and touched off a race in the tech world to bring AI into the mainstream.

Since then, computing's AI age has been taking shape largely behind the scenes. Machine learning, an underlying technology that involves computers learning from data, has been widely used in jobs such as identifying credit card fraud and making online content and advertising more relevant. If the robots are starting to take all the jobs, it's been happening largely out of sight.

That is, until now. Another breakthrough in AI has just shaken up the tech world. This time, the machines are operating in plain sight — and they could finally be ready to follow through on the threat to replace millions of jobs.

ChatGPT, a query-answering and text-generating system released at the end of November, has burst into the public consciousness in a way seldom seen outside the realm of science fiction. Created by San Francisco-based research firm OpenAI, it is the most visible of a new wave of so-called "generative" AI systems that can produce content to order.

If you type a query into ChatGPT, it will respond with a short paragraph laying out the answer and some context. Ask it who won the 2020 US presidential election, for example, and it lays out the results and tells you when Joe Biden was inaugurated.

Simple to use and able in an instant to come up with results that look like they were produced by a human, ChatGPT promises to thrust AI into everyday life. The news that Microsoft has made a multibillion dollar investment in OpenAI — co-founded by AlexNet creator Sutskever — has all but confirmed the central role the technology will play in the next phase of the AI revolution.

ChatGPT is the latest in a line of increasingly dramatic public demonstrations. Another OpenAI system, automatic writing system GPT-3, electrified the tech world when it was unveiled in the middle of 2020. So-called large language models from other companies followed, before the field branched out last year into image generation with systems such as OpenAI's Dall-E 2, the open-source Stable Diffusion from Stability AI, and Midjourney.

These breakthroughs have touched off a scramble to find new applications for the technology. Alexandr Wang, chief executive of data platform Scale AI, calls it "a Cambrian explosion of use cases", comparing it to the prehistoric moment when modern animal life began to flourish.

If computers can write and create images, is there anything, when trained on the right data, that they couldn't produce? Google has already shown off two experimental systems that can generate video from a simple prompt, as well as one that can answer mathematical problems. Companies such as Stability AI have applied the technique to music.

The technology can also be used to suggest new lines of code, or even whole programs, to software developers. Pharmaceutical companies dream of using it to generate ideas for new drugs in a more targeted way. Biotech company Absci said this month it had designed new antibodies using AI, something it said could cut more than two years from the approximately four it takes to get a drug into clinical trials.

But as the tech industry races to foist this new technology on a global audience, there are potentially far-reaching social effects to consider.

Tell ChatGPT to write an essay on the Battle of Waterloo in the style of a 12-year-old, for example, and you've got a schoolchild's homework delivered on demand. More seriously, the AI has the potential to be deliberately used to generate large volumes of misinformation, and it could automate away a large number of jobs that go far beyond the types of creative work that are most obviously in the line of fire.

'It's a tool that people can use morally or immorally, legally or illegally, ethically or unethically'

"These models are going to change the way that people interact with computers," says Eric Boyd, head of AI platforms at Microsoft. They will "understand your intent in a way that hasn't been possible before and translate that to computer actions". As a result, he adds, this will become a foundational technology, "touching almost everything that's out there".

The reliability problem

Generative AI advocates say the systems can make workers more productive and more creative. A code-generating system from Microsoft's GitHub division is already coming up with 40 per cent of the code produced by software developers who use the system, according to the company.

The output of systems like these can be "mind unblocking" for anyone who needs to come up with new ideas in their work, says James Manyika, a senior vice-president at Google who looks at technology's impact on society. Built into everyday software tools, they could they suggest ideas, check work and even produce large volumes of content.

Yet for all its ease of use and potential to disrupt large parts of the tech landscape, generative AI presents profound challenges for the companies building it and trying to apply it in practice, as well as for the many people who are likely to come across it before long in their work or personal lives.

Foremost is the reliability problem. The computers may come up with believable-sounding answers, but it's impossible to completely trust anything they say. They make their best guess based on probabilistic assumptions informed by studying mountains of data, with no real understanding of what they produce.

"They don't have any memory outside of a single conversation, they can't get to know you and they don't have any notion of what words signify in the real world," says Melanie Mitchell, a professor at the Santa Fe Institute. Merely churning out persuasive-sounding answers in response to any prompt, they are brilliant but brainless mimics, with no guarantee that their output is anything more than a digital hallucination.

There have already been graphic demonstrations of how the technology can produce believable-sounding but untrustworthy results.

Late last year, for instance, Facebook parent Meta showed off a generative system called Galactica that was trained on academic papers. The system was quickly found to be spewing out believable-sounding but fake research on request, leading Facebook to withdraw the system days later.

ChatGPT's creators admit the shortcomings. The system sometimes comes up with "nonsensical" answers because, when it comes to training the AI, "there's currently no source of truth", OpenAI said. Using humans to train it directly, rather than letting it learn by itself — a method known as supervised learning — did not work because the system was often better at finding "the ideal answer" than its human teachers, OpenAI added.

Microsoft has made a multibillion dollar investment in research outfit OpenAI, which created ChatGPT

One potential solution is to submit the results of generative systems to a sense check before they are released. Google's experimental LaMDA system, which was announced in 2021, comes up with about 20 different responses to each prompt and then assesses each of these for "safety, toxicity and groundedness", says Manyika. "We make a call to search to see, is this even real?"

Yet any system that relies on humans to validate the output of the AI throws up its own problems, says Percy Liang, an associate professor of computer science at Stanford University. It might teach the AI how to "generate deceptive but believable things that actually fool humans," he says. "The fact that truth is so slippery, and humans are not terribly good at it, is potentially concerning."

According to advocates of the technology, there are practical ways to use it without trying to answer these deeper philosophical questions. Like an internet search engine, which can throw up misinformation as well as useful results, people will work out how to get the most out of the systems, says Oren Etzioni, an adviser and board member at A12, the AI research institute set up by Microsoft co-founder Paul Allen.

"I think consumers will just learn to use these tools to their benefit. I just hope that doesn't involve kids cheating in school," he says.

But leaving it to the humans to second-guess the machines may not always be the answer. The use of machine-learning systems in professional settings has already shown that people "over-trust the predictions that come out of AI systems and models", says Rebecca Finlay, chief executive of the Partnership on AI, a tech industry group that studies uses of AI.

The problem, she adds, is that people have a tendency to "imbue different aspects of what it means to be human when we interact with these models", meaning that they forget the systems have no real "understanding" of what they are saying.

These issues of trust and reliability open up the potential for misuse by bad actors. For anyone deliberately trying to mislead, the machines could become misinformation factories, capable of producing large volumes of content to flood social media and other channels. Trained on the right examples, they might also imitate the writing style or spoken voice of particular people. "It's going to be extremely easy, cheap and broad-based to create fake content," says Etzioni.

'It's going to be extremely easy, cheap and broad-based to create fake content'

This is a problem inherent with AI in general, says Emad Mostaque, head of Stability AI. "It's a tool that people can use morally or immorally, legally or illegally, ethically or unethically," he says. "The bad guys already have advanced artificial intelligence." The only defence, he claims, is to spread the technology as widely as possible and make it open to all.

That is a controversial prescription among AI experts, many of whom argue for limiting access to the underlying technology. Microsoft's Boyd says the company "works with our customers to understand their use cases to make sure that the AI really is a responsible use for that scenario".

He adds that the software company also works to prevent people from "trying to trick the model and doing something that we wouldn't really want to see". Microsoft provides its customers with tools to scan the output of the AI systems for offensive content or particular terms they want to block. It learnt the hard way that chatbots can go rogue: its Tay bot had to be hastily withdrawn in 2016 after spouting racism and other inflammatory responses.

To some extent, technology itself may help to control misuse of the new AI systems. Manyika, for instance, says that Google has developed a language system that can detect with 99 per cent accuracy when speech has been produced synthetically. None of its research models will generate the image of a real person, he adds, limiting the potential for the creation of so-called deep fakes.

Jobs under threat

The rise of generative AI has also touched off the latest round in the long-running debate over the impact of AI and automation on jobs. Will the machines replace workers or, by taking over the routine parts of a job, will they make existing workers more productive and increase their sense of fulfilment?

Most obviously, jobs that involve an substantial element of design or writing are at risk. When Stability Diffusion appeared late last summer, its promise of instant imagery to match any prompt sent a shiver through the commercial art and design worlds.

Some tech companies are already trying to apply the technology to advertising, including Scale AI, which has trained an AI model on advertising images. That could make it possible to produce professional-looking images from products sold by "smaller retailers and brands that are priced out of doing photoshoots for their goods," says Wang.

That potentially threatens the livelihoods of anyone who creates content of any kind. "It revolutionises the entire media industry," says Mostaque. "Every single major content provider in the world thought they needed a metaverse strategy: they all need a generative media strategy."

According to some of the humans at risk of being displaced, there is more at stake than just a pay cheque. Presented with songs written by ChatGPT to sound like his own work, singer and songwriter Nick Cave was aghast. "Songs arise out of suffering, by which I mean they are predicated upon the complex, internal human struggle of creation and, well, as far as I know, algorithms don't feel," he wrote online. "Data doesn't suffer."

Techno-optimists believe the technology could amplify, rather than replace, human creativity. Armed with an AI image generator, a designer could become "more ambitious", says Liang at Stanford. "Instead of creating just single images, you could create whole videos or whole new collections."

The copyright system could end up playing an important role. The companies applying the technology claim that they are free to train their systems on all available data thanks to "fair use", the legal exception in the US that allows limited use of copyrighted material.

Others disagree. In the first legal proceedings to challenge the AI companies' profligate use of copyrighted images to train their systems, Getty Images and three artists last week started actions in the US and UK against Stability AI and other companies.

According to a lawyer who represents two AI companies, everyone in the field has been braced for the inevitable lawsuits that will set the ground rules. The battle over the role of data in training AI could become as important to the tech industry as the patent wars at the dawn of the smartphone era.

Ultimately, it will take the courts to set the terms for the new era of AI — or even legislators, if they decide the technology breaks the old assumptions on which existing copyright law is based.

Until then, as the computers race to suck up more of the world's data, it is open season in the world of generative AI.