WORK AND DIGITALISATION 26.01.2023 | Barry Eichengreen

Sympathy for the algorithm

The release of ChatGPT garnered great attention and raised even greater questions. How worried should we be?

With hindsight, 2022 will be seen as the year when artificial intelligence gained street credibility. The release of ChatGPT by the San Francisco-based research laboratory OpenAI garnered great attention and raised even greater questions.

In just its first week, ChatGPT attracted more than a million users and was used to write computer programs, compose music, play games and take the bar exam. Students discovered that it could write serviceable essays worthy of a B grade – as did teachers, albeit more slowly and to their considerable dismay.

ChatGPT is far from perfect, much as B-quality student essays are far from perfect. The information it provides is only as reliable as the information available to it, which comes from the internet. How it uses that information depends on its training, which involves supervised learning, or, put another way, questions asked and answered by humans.

An obvious implication is more rapid displacement from jobs, compared to past waves of automation and more rapid restructuring of surviving jobs.

The weights that ChatGPT attaches to its possible answers are derived from reinforcement learning, where humans rate the response. ChatGPT's millions of users are asked to upvote or downvote the bot's responses each time they ask a question. In the same way useful feedback from an instructor can sometimes teach a B-quality student to write an A-quality essay, it's not impossible that ChatGPT will eventually get better grades.

This rudimentary artificial intelligence forces us to rethink what tasks can be carried out with minimal human intervention. If an AI is capable of passing the bar exam, is there any reason it can't write a legal brief or give sound legal advice? If an AI can pass my wife's medical-licensing exam, is there any reason it can't provide a diagnosis or offer sound medical advice?

An obvious implication is more rapid displacement from jobs, compared to past waves of automation and more rapid restructuring of surviving jobs. And the jobs that will be automated out of existence will not be limited to the low-skilled and low-paid.

Less obvious is who is safe from technological unemployment. What human traits, if any, will an AI be unable to simulate? Are those traits innate, or can they be taught?

Teaching empathy

The safest jobs will be those requiring empathy and originality. Empathy is the ability to understand and share the feelings and emotions of others. It creates the interpersonal compassion and understanding that are fundamental to social interactions and emotional well-being. It is especially valuable in circumstances and periods of difficulty. That's why empathy is valued in religious leaders, caregivers and grief counsellors.

It is possible to imagine that, with the help of facial-recognition software, an AI can learn to *recognise* the feelings of its interlocutors (that it can learn what is known as 'cognitive empathy'). But it can't obviously *share* their feelings (it can't learn 'affective empathy') in the same way that my wife, in her empathic moments, shares my feelings. Add that to the list of reasons why an AI can't replace my wife, my doctor, or my rabbi.

There is no consensus about whether affective empathy can be cultivated and taught. Some argue that affective empathy is triggered by mirror neurons in the brain that can't be artificially stimulated or controlled. Empathy is just something we experience, not something we can learn. It follows that some of us are better wired than others to be caregivers and grief counsellors.

Other researchers suggest that this emotional response can indeed be taught. There is even a training company for medical clinicians called Empathetics, Inc. If true, it may be possible that more people can be prepared for automation-safe jobs where affective empathy is required.

But if humans can learn affective empathy, then why can't algorithms? The idea that jobs requiring affective empathy will remain safe from automation assumes that people can distinguish true empathy from the simulation.

Originality means doing something that hasn't

product, DALL•E, is able to generate sophisticated images from text descriptions. This has created some consternation among artists.

been done previously, for example, creating a painting, composition or newspaper commentary wholly unlike what has come before. Originality is distinct from creativity, which involves combining pre-existing elements in novel ways.

Another OpenAI product, DALL•E, is able to generate sophisticated images from text descriptions ('a painting of an apple' or 'the "Mona Lisa" with a moustache'). This has created some consternation among artists. But are its responses, derived using a large dataset of text-and-image pairs, original artwork?

It is questionable whether they are original in the sense of portraying an aesthetically pleasing image unlike any seen before, as opposed to combining existing visual elements associated with existing text. Artists who trade on originality may have nothing to fear, assuming of course that viewers can distinguish original artwork from the rest.

Again, there is no consensus on whether originality is inborn or can be taught. The answer, most likely, is: a bit of both.

How worried should we be? Type 'Write an 800-word commentary on AI for *Project Syndicate*' into ChatGPT and judge for yourself.

©☒ Project Syndicat



Barry Eichengreen

Barry Eichengreen is a Professor of Economics at the University of California as well as a former senior policy adviser at the International Monetary Fund. He is the author of many books, including *In Defense of Public Debt* (Oxford University Press, 2021).