

The ChatGPT revolution of academic research has begun

The AI chatbot may soon kill the undergraduate essay, but its transformation of research could be equally seismic. Jack Grove examines how ChatGPT is already disrupting scholarly practices and where the technology may eventually take researchers – for good or ill



Source: Getty/iStock montage

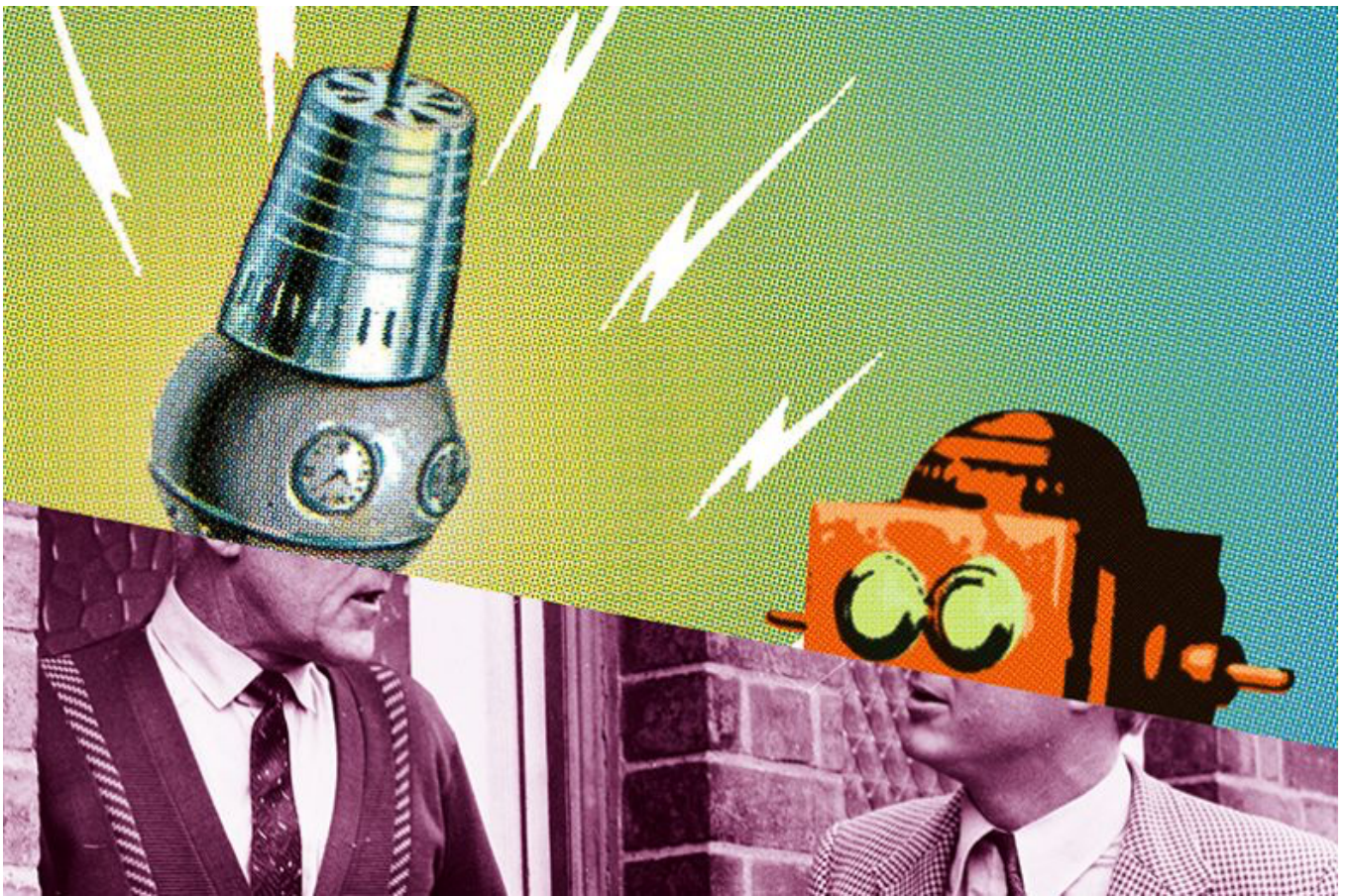
For politicians, pundits and psephologists, the tens of thousands of open-text comments regularly collected by the [British Election Study](#) (BES) offer a unique and valuable insight into the minds of the UK public. However, the research assistants required to trawl this ocean of text and sort the responses – some 657,000 since 2014 alone – into categories could be forgiven for feeling less enthusiastic about the project. But maybe for not much longer. According to a [preprint](#) published in December, ChatGPT, the “[large language model](#)” (LLM) chatbot launched in December, was able to code responses with a 92 per cent accuracy rate compared with a trained human coder.

“Coding responses is very time-consuming. I’ve done a few shifts of it and you can do 2,000 to 3,000 responses in a day,” says Ralph Scott, a

research associate on the BES based at the [University of Manchester](#), who co-authored the ChatGPT study with colleagues at Manchester and West Point Military Academy. Aided by artificial intelligence, the past three internet surveys, covering more than 81,000 people, were sifted within seconds, says Scott. "It makes those parts of research which don't need creativity or judgement so much easier," he explains.

Freeing up time and money for higher-level analysis are not the only immediate benefits that ChatGPT offers, in Scott's view. It also opens new avenues for social science inquiry, he believes. This is because even BES researchers, backed by multimillion-pound research grants, must still be cautious about setting too many open-text questions for fear of being flooded with data that will take time and money to process, he says. "What is exciting with ChatGPT is that you can ask a lot more questions because we analyse the data so much more easily," he says. These extra questions "could reveal there are whole undercurrents in political thought that we just haven't considered".

Indeed, it is conceivable that standard opinion polling, with questions based around the five-point Likert scale, may quickly become obsolete now that richer textual data can be scanned and structured in seconds.



ChatGPT's launch has generated almost as many column inches as the technology itself has produced in response to prompts by millions of fascinated users. Many of those column inches have been in the academic press, as university teachers in their droves have agonised over what the chatbot's startling essay-writing ability means for the future of university [teaching and assessment](#) – in good ways and bad. However, the technology's implications for university research – beyond the obvious potential to [scrape together an anodyne publication](#) for an unselective journal – have been less widely discussed. Yet they may be just as profound.

"Every so often, you wake up and realise we are in the future," reflects Toby Walsh, professor of artificial intelligence at [UNSW Sydney](#) and Australian Laureate Fellow. In terms of "what's under the bonnet", ChatGPT is "not very different from what has been about for some time, but they've done an exceptional job of making it user-friendly", says Walsh. However, while general users have enjoyed the technology's limerick-writing ability, the "big problem" for academics is that LLMs "can just make stuff up", Walsh warns.

For example, scholars who asked the technology to list the top papers in their fields have been offered entirely fictional lists of titles and authors. Others have been presented with invented versions of their own CVs. One of those was Walsh.

“Everything that [ChatGPT] wrote was plausible but completely made up – except that I was happily married!” he notes. “It decided that I’d given up academia to compete as a professional poker player who’d won the World Series. Some people in my walk of life have done this, so it decided to invent this story [in my case].”

The technology’s reliance on language-scraping from its training materials is why the expected [launch of ChatGPT-4](#) later this year might not be the game-changer that some expect, says Walsh, even though its training database is 500 times larger than ChatGPT-3’s. “Because [ChatGPT] cannot reason, it will still say stupid things,” he says. “For instance, if you explain that Jim has three red balloons and Sue has nine blue balloons and then four pop, it won’t give the right answer [about how many are left],” he explains. “It will find the most common answer on the internet regarding red balloons and simply say what is probable.”

In Walsh’s view, the trick to making ChatGPT a more powerful tool would be to give researchers the ability to restrict the data sources it harvests, rather than adding more. “If you could curate what it looks at – say, restrict it to the first five links on Google – that would be much more useful, but I think that’s a year or two away,” he says.

“If you could ask ChatGPT to focus only on papers with a minimum of three citations, published in major journals since the 1950s, then that could be a super-powerful tool for academics,” agrees [Andy Farnell](#), visiting and associate professor of signals, systems and cybersecurity at a number of European universities. Alternatively, a sophisticated search of a digitised university archive might surface links or patterns in the study of history, literature or other fields that might have been overlooked for decades. Much of these records are, however, still inaccessible to ChatGPT, he says.

"This raises the question of whether universities become consumers of this technology at the mercy of Big Tech, or whether they seek to bring it in-house by creating their own versions – which places like MIT and Berkeley may feel they want to do," says Farnell. "If you're an institution that can fund your own AI and compute around your own intellectual property, that might give you a huge advantage – it could even redefine what it means to be a university. If you're UCL, would you want to open up your database for others to make these breakthroughs, or would you want to do things yourself?"

It is, however, debatable how many universities would have the finances to develop their own large language algorithm, Farnell concedes. "There are huge obstacles to creating your own AI – not least the costs of power consumption needed to create and run these supercomputers. They require the same amount of energy as it takes to manufacture a 747 aircraft," says Farnell, who believes the "politics of how AI is distributed may become quite divisive".

But, if nothing else, ChatGPT has underlined to everyone that AI is a technology that cannot be ignored, Farnell says: "In some ways, the big leap forward isn't so much the technology but the acceptance that AI will change things. It means that AI will find a purchase in places where it hasn't been seen before.

That is why "in the same week that [Microsoft fired 10,000 staff](#)" in mid-January, the Silicon Valley giant decided to invest £10 billion in ChatGPT's parent company, OpenAI. "That speaks to the [firm's] confidence about where things are heading," he says. Microsoft is reportedly keen to incorporate ChatGPT's technology into its search engine, Bing, which has been a very poor relation to Google over the past decade. The technological advances will only accelerate as Google, [aware of the threat](#), fine-tunes its own LLM, Farnell adds. With Google's skunkworks AI project Deep Mind already making waves in academic research, from its protein-structure [prediction technology](#) announced in 2020 to its work decrypting ancient texts via its [Ithaca](#) neural network, what Google will

produce will be “very cool”, says Farnell – although \$100 billion was knocked off the value of Google’s parent company, Alphabet, when its LLM, known as Bard, [gave inaccurate information](#) during its first demonstration in early February.

Regardless, researchers will need to think quickly about how to interact with ChatGPT and its rivals in their teaching and research. For example, Farnell “cannot see myself being able to use the same research methods materials for PhD students as I did last year”.

Some researchers have experimented with using the technology to help them complete the more tedious aspects of grant applications. “If you want a starting draft for some sections on applications – say, the potential academic benefits or the communication plan – [ChatGPT] is quite good,” says John Tregoning, reader in respiratory infections at [Imperial College London](#). “It also writes plausible references.”

Walsh, meanwhile, suggests that the technology could also ease the pain of university bureaucracy. “It could easily set your key performance indicators – ‘I’m going to do this amount of research, teaching and professional development’, he says, while others claim it works wonders for [reformatting citations and footnotes](#) to the various styles required by different journals.

Some scholars may be tempted to use ChatGPT to help write papers, too. For their part, Scott and colleagues used it to write the abstract of their paper about the technology’s ability to code survey responses. Enlisting ChatGPT’s help to write literature review sections or even whole review papers has also been mooted, but, for Tregoning, this appears a bridge too far at this stage. “I asked it to write a student essay and it produced [something that would get] a 2:2 at undergraduate level but would fail at master’s level. The level of knowledge synthesis wasn’t there. It can’t analyse or evaluate,” he says, echoing Walsh’s point about the technology’s inability to reason.

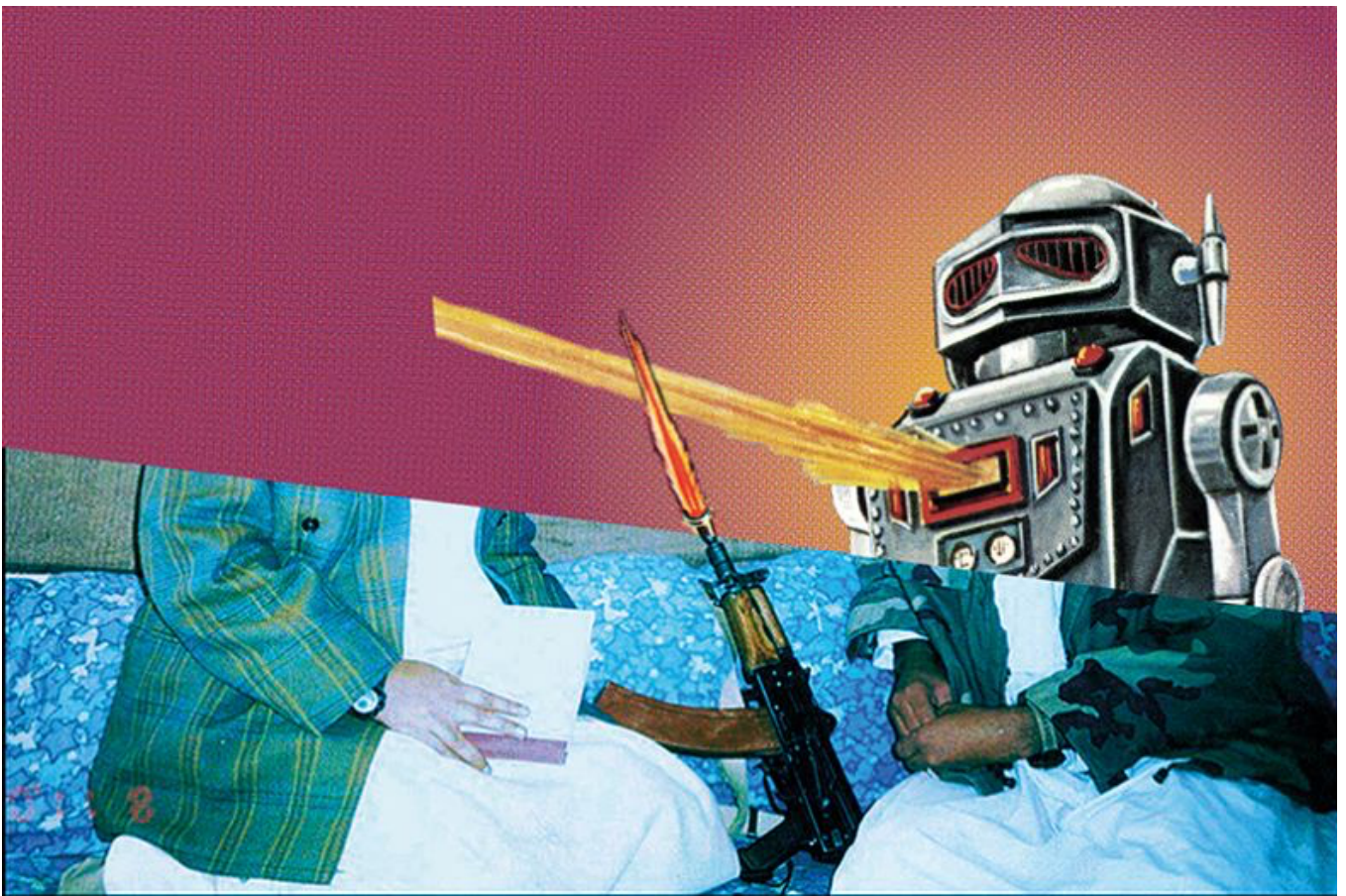
But others are less downbeat. For instance, [Roger Watson](#), editor-in-

chief of *Nurse Education in Practice*, recently used SciSpace – another AI-powered writing tool – to summarise a very complicated abstract. The results, he said, were “amazing”, and he is now working on a project to compare whether AI-generated literature reviews measure up to those written by humans.

But how far should robot-writing go? Do AI-written papers pose the same threat to research integrity as AI-written essays are widely considered to pose to academic integrity in student assessment?

Several preprints and published articles have [already credited ChatGPT with formal authorship](#), but a January editorial in *Nature* took issue with this practice on the ground that “attribution of authorship carries with it accountability for the work, and AI tools cannot take such responsibility”. Instead, the journal requires researchers using LLM tools to “document this use in the methods or acknowledgements sections” of a paper.

But what if researchers, instead, decide to take all the credit for themselves for an AI-authored paper? Will they be found out? *Nature's* publisher, Springer Nature, is among those developing technologies to detect text generated by LLMs. The *Nature* editorial notes that the robots’ current writing “can appear bland and generic, or contain simple errors. Moreover, they cannot yet cite sources to document their outputs.” However, it adds, “LLMs will improve, and quickly”, such as by being linked to source-citing tools. In the end, “researchers should ask themselves how the transparency and trust-worthiness that the process of generating knowledge relies on can be maintained if they or their colleagues use software that works in a fundamentally opaque manner.”



Some AI technologies have already demonstrated an ability – in tandem with humans – to produce fluent papers that are also scientifically sound and insightful. Casey Greene, founding director of the University of Colorado’s Center for Health AI, recently published a [preprint](#) with Milton Pivodori, from the [University of Pennsylvania](#), showing how the AI writing tool Manubot is able to sharpen up almost every paragraph of a journal article by summarising scientific concepts and producing “high-quality revisions that improve clarity”. Such technological advances will “revolutionise the type of knowledge work performed by academics” – with scholars able to devote more time to experiments and fewer hours to finessing final drafts, the paper predicts.

“Scientific manuscripts are written according to somewhat [arcane rules](#) and not having to live and breathe those rules can help us focus on what we’re communicating, as opposed to how,” Greene tells *Times Higher Education*.

AI tools could lead to “a more equitable and productive future for research, where scientists are only limited by their ideas and ability to

conduct experiments", the paper adds, hinting at a world where researchers in non-English speaking countries will not be hampered by language barriers when sending their papers to top English-language journals. "This would be the ideal outcome, and I think it is a possible one," says Greene, who notes that universities with the resources to pay for life sciences editors – such as his own institution – also have a head start on those in the US that don't. "If this technology can be deployed to cross these gaps, it can decrease the unwritten costs of participation in research," says Greene.

AI tools might also allow scholars to publish outside their field or in a genuinely interdisciplinary way, Greene adds. "I was trained in genetics and when our work ends up adjacent to immunology, I always struggle to remember the various clusters of differentiation, perhaps because I think in protein/gene names," he says. "Maybe the next version [of AI] can transform those manuscripts into terms that I can recall more quickly, reducing the cognitive load."

Greene admits that the current iteration of Manubot is not always accurate either. "I imagined this as an academic writing partner, but one of the things that surprised me was that, at some points in our manuscript, the language model described steps that we had not taken," reflects Greene. But one of these steps might actually turn out to be a valuable addition, he says, raising the possibility that AI might become able to offer useful prompts for additional experiments. "There are hints in our work that this might also be possible," Greene believes.

That may lead to some fortunate breakthroughs – although the potential for misuse of such technology beyond the scientific arena is also obvious. "Some of its potential is very scary," reflects Farnell. "You could ask it to imagine 1,000 ways to commit a terrorist plot. Even if just one of those ideas worked, or inspired someone to do something awful, then that's a terrible thing. People are already using it to create malware – though the potential to cause harm is probably more contained if it's being used within an institution," he says.

University researchers will also need to beware of the biases baked into LLMs because the internet content on which they are trained has largely been created by young white men. "If you have overly hegemonic trading data, this will encode a system of oppression and biases" into the technology, warned Emily Bender, a [University of Washington](#) computational linguist, in an [Alan Turing Institute lecture](#) in 2021, pointing to a [Massachusetts Institute of Technology study](#) in 2020 that identified bias in nearly all natural language processing systems.

Perhaps unsurprisingly, ChatGPT itself, when asked what benefits it offers researchers, does not recognise such biases. Rather, it states that it can "eliminate errors caused by human biases or oversights, providing a more objective and accurate perspective on research questions". It also argues that it can "sort and organise information more effectively than humans". Such declarations might be thought to support Farnell's view that "talking to ChatGPT is like talking to a psychopath: it has no sense of fallibility and gives you just enough of what you want to hear".

Yet the chatbot is not so immodest as all that. It goes on: "I don't believe I can make human researchers unnecessary. Human researchers bring unique skills and perspectives that are critical to advancing scientific understanding. They are better at generating new research questions, designing experiments, and interpreting complex data. While AI can assist with many aspects of research, it cannot replace the creativity and critical thinking skills of human researchers. Instead, AI can be seen as a tool to augment human researchers' capabilities, making research faster, more efficient and more accurate."

If these claims are borne out by the early experimenters, it is hard to imagine researchers turning down such mechanical help in the highly competitive race to publication, even if it is being offered by a psychopath. They are only human, after all.

jack.grove@timeshighereducation.com

"There are smart ways to use ChatGPT and there are dumb ways"

As a Fulbright scholar living in upstate New York, Mushtaq Bilal used to talk to himself on his way to his daily swim. "As I drove to the pool, I'd have some good ideas about my next paper, so thought I should record and transcribe them," recalls the Pakistani academic, now a postdoctoral scholar in world literature at the [University of Southern Denmark](#). "That would give me a zero-draft, which I could work on."

That method has, however, become much easier with ChatGPT, says Bilal, whose tips on how to use the AI tool for academic writing have been viewed [millions of times](#) on Twitter, pushing his following past 125,000. "You can tell it to remove redundant words, create coherent sentences and cohesive paragraphs," he tells *Times Higher Education*. This method can [compress a block of 13,000 words](#) – the result of two hours of dictation – into a presentable 3,000-word essay, which, with another two hours of scholarly fine-tuning, is ready for submission.

Bilal rejects the suggestion that using AI is cheating. After all, the ideas are his own and the final draft will require many small edits to pass muster. "I'm a big believer in making notes by hand and I read very slowly, but I recognise that getting that first draft can be difficult," he says.

To strike the right tone, Bilal imagines he is talking to a "very keen, intelligent first-year undergraduate", even when instructing ChatGPT to improve text. "I talk to it politely...I say 'please do this,' even though it's a machine. You need to have a positive relationship with it."

With ChatGPT and other AI language tools here to stay, the challenge is how to use them intelligently and ethically, continues Bilal. "There are smart ways to use ChatGPT and there are dumb ways – one of which is asking it questions," he says.

"Asking for answers is the wrong approach – you need to ask it to give

you questions, so you can create your own answers," he explains. For example, he recently asked ChatGPT to imagine 10 questions that might be asked in a Fulbright interview, reflecting on how [he was initially rejected](#) for the prestigious US scholarship before securing a PhD place at Binghamton University. "Of these 10 questions, I was asked eight," he reflects.

Having these tools will help to level the playing field for those unacquainted with the unspoken rules that govern many parts of academia, says Bilal. "If you want to prepare an application for a US graduate school, there are things like a statement of purpose that I had no knowledge of," he explains. "It took me 12 to 18 months to find the information and fill in an application, but you could easily ask ChatGPT to write a statement for you. It would be foolish just to submit this because it would have no personality, but it would give you the right structure on which you could build."

From writing considerate rejection letters to more punchy Twitter threads (he recommends WordTune), Bilal has found numerous ways to employ AI in academic life. While he acknowledges it might not suit everyone, academics should be open to its benefits: "If you are willing learner – as most researchers are – why wouldn't you want to learn this craft?"