CURRENT SCIENCE

Volume 125 Number 9 10 November 2023

GUEST EDITORIAL

Good science writing in the days of AI and ChatGPT

For most researchers, writing is a chore, a task viewed as a necessary evil, undertaken with reluctance, executed with indifference and completed with relief. It is no surprise that researchers have welcomed tools like ChatGPT while journals have accepted, if not endorsed, their use. Copy editors, however, have had some reservations about their effectiveness. The few texts that I have come across written with the help of AI – some with acknowledgement of that fact and some without – were unexceptional, with impeccable grammar, wide vocabulary, and convoluted prose that looked as though indeed written by veteran academics – and with as much originality as one may expect from a photocopier. That is what I wish to take as my point of departure, namely that AI tools can generate serviceable prose but the output is best not submitted without scrutiny and that researchers, at least early-career researchers, would do well to become more competent writers.

The text generated by AI requires scrutiny because GPT stands for 'generative pre-training transformer'. Essentially, the tool is a prediction engine: given a string of words, it predicts the most probable next word; given a string of sentences, it predicts the most probable next sentences. Because these probabilities are derived from massive data sets, truly original stretches of text are generated rarely – making it likely that such text will score highly on 'similarity' if checked for plagiarism using tools like iThenticate and Unicheck. Secondly, please bear in mind that the tools generate the *most probable* text, not necessarily *true* text. On top of that, the tool is known to fabricate citations and references – which means that your research paper may contain non-existent references. And that will not show you in a good light should a diligent reviewer decide to look up each reference. On the other hand, many publishers do not prohibit the use of AI but ask that authors mention up front, in the cover letter, if they used such a tool and, if they did, which tool (name and version) and to what end, and also state that full responsibility for the content rests with the authors, not with the tool.

Then there is the security angle. All inputs into any such tool and its output are stored and thus open to probing and even misuse by whosoever has access to such stored texts. If your research is sensitive or potentially exploitable commercially, think twice before using a draft of your research paper as input.

Of these possible pitfalls, let me focus on originality. Meeting the demands for originality in research is hard enough, and most researchers would rather do without it in their writing. I am tempted to quote C. Northcote Parkinson here, the originator of the famous Parkinson's Law, from his short essay on why journals proliferate, in which he says that, because of rivalry between academics, Professor B 'will accept articles from all who are not actual and known adherents of A (the editor of a rival journal). He eventually draws a line, however, at contributions from C, whose works are confused, long, and *original only in their grammar and punctuation*' (emphasis added). Without going that far in our desire to be original, I wish researchers would put in at least a modicum of effort to make their writing clear, concise, and most important, reader-friendly. If any incentive is needed, it lies in the observation that well-written papers are not only more likely to be accepted but also cited more often than poorly written papers. Such writing is possible but takes some systematic effort and investment in time spent on reading (or even listening to) well-written books. I say books because the text is generally better than that encountered in newspapers, magazines and even journals.

But what is good writing? Although that master of competent writing, William Somerset Maugham, suggested that good prose is simple, lucid and euphonious, he did not have to publish research papers, a genre quite unlike short stories, novels, plays and even essays – the genres in which Maugham excelled. And yet, good science writing is not hard to find. Here is one example which makes the point that a signal travelling along a neural network and crossing the 'synaptic cleft' – essentially a void or a break in the network with no physical continuity – represents a process: 'Like lips whispering very close to an ear, synaptic communication between neurons has three basic components: the presynaptic terminal of the axon, which sends signals (corresponding to lips); the synaptic cleft (the space between lips and ears); and the post-synaptic site on the dendrite that receives signals (the ear).' This is Eric Kandel – who won the Nobel Prize for his work on memory – writing for the general reader.

Here is another example which shows the impact of being specific by bringing in numbers or quantities. Compare the following two short pieces of prose: (1) 'Pitchblende contains uranium. Marie and Pierre Curie purified huge amounts of the natural ore.' (2) 'From several tonnes of pitchblende, 400 tonnes of washing water, and hundreds of buckets of distilled sludge waste, they extracted 0.1 g of radium.' The second piece has greater force because it is specific and concrete – and shows why its author, Siddhartha Mukherjee, is praised for the quality of his writing.

Here is a bit longer example, which shows the impact of little details that paint pictures in readers' minds. Many readers may have seen a creeper that manages to cover even bare walls with its foliage. How does it manage to cling to such inhospitable surfaces? Not by penetrating any crevices between bricks or stones with its roots, as is commonly believed, but by secreting 'minute drops of clear fluid, not in the least milky like that exuded from a wound. This fluid is slightly viscid but cannot be drawn out into threads. It has the remarkable property of not soon drying … rootlets were left in contact with the glass for about ten days or a fortnight, and the drops of secreted fluid were now rather larger and so viscid that they could be drawn out into threads … Hence, we may conclude that the rootlets first secrete a slightly viscid fluid, subsequently absorb the watery parts (for we have seen that the fluid will not dry by itself) and ultimately leave a cement.' This is a simple explanation of the mechanism, presented clearly and vividly and with no jargon – by Charles Darwin in 1876.

Models of good writing are not hard to find, provided one knows where to look for them. Anthologies offer a good place to start: *The Oxford Book of Modern Science Writing*, compiled by Richard Dawkins; *From Creation to Chaos: Classic Writings in Science*, by Bernard Dixon (a former editor of *New Scientist*); and *Best Science Writing: Readings and Insights*, by Robert Grannon. The Royal Society has been awarding a prize every year since 1988 to the best-written science book for the general public; the winner, as well as the books shortlisted for the prize each year, make great reading and span a wide variety of subjects. The US equivalent of such a prize is the PEN/E O Wilson Literary Science Writing Award.

Over the years, I have been practising what I preach and have been able to identify some of the strategies that award-winning writers deploy.

- Establish context or set the stage before plunging into details.
- Use concrete terms and supply specific details (of shape, size, weight, colour, texture or whatever) to help readers visualize – see in their mind's eye – the objects or actions being described.
- Give examples.
- Develop analogies to connect what readers already know to what is new.
- Ask questions and then answer them.
- Repeat the key concepts or important points but in different words.
- Use personal pronouns (I, we, you) and address your readers directly to give a conversational tone to your writing.

Do these strategies really work? Yes, they do. And if you wish to know why and how in terms of what happens within your brain – as backed up by research – I suggest you read *Writing for Impact* by Bill Birchard. For example, he quotes a research paper that showed that specific words, or words 'for which the visual form is better specified', are more richly encoded in the brain than words 'denoting basic level category members lacking a specific form' – in other words, 'a crow' is more specific than 'a bird'.

But readers of *Current Science* are researchers and academics; they demand hard evidence. Which is what Helen Sword supplies in *Stylish Academic Writing*. She analysed in detail a representative corpus of academic writing comprising 500 research papers, 50 from each of the 10 domains or disciplines ranging from medicine, computer science and evolutionary biology to law, philosophy and literary studies. The analysis focused on such stylistic features as the use of personal pronouns, the proportion of abstract nouns, and engaging titles and openings. Based on this and other work, she has developed an algorithm that can test samples of any writing. The results of this 'test' are presented as a terse medical report. Paste a sample of your writing (100 to 1000 words long) in the box provided at https://writers-diet.com/writing-test/ and see what the report says: maybe it will say 'Fit & trim: in excellent condition' or 'Needs toning: would benefit from a light workout' or 'Flabby: judicious editing required'. I fed a sample of text from a draft of this guest editorial and was rewarded with 'Fit & Trim'. Now it is your turn.

Yatheendra Joshi

A-1/702 Land-mark Garden,
78 Kalyani Nagar,
Pune 411 006, India
e-mail: yateendra.joshi@gmail.com