

Reflective robots - The impact of AI on medical writing

Luke Nugent



Luke Nugent is an ST8 trainee in the West Midlands' Stoke/Oswestry rotation. His interest is hand surgery and he is a BSSH diplomat, but has also written for the speleology publication *'Descent'* on the subject of AI's utility in mapping 18th century mine works in Derbyshire.

Surgeons, trainees, and the public in general have in the past couple of years been very rapidly exposed to the everyday use of artificial intelligence (AI). Inevitably, this has significantly impacted various aspects of our lives, and healthcare and medical research are no exceptions. Generative AI, a subset of AI focused on producing new content from vast datasets, has become a powerful tool in medical literature and clinical practice. Generative AI encompasses widely available models such as OpenAI's GPT-4 (Generative Pre-trained Transformer 4) and Google's BERT (Bidirectional Encoder Representations from Transformers), which have evolved from simple algorithms to complex systems capable of understanding and generating human-like text^{1,2}. These models work by training on massive amounts of text data,

learning patterns, structures, and contextual cues, enabling them to generate coherent and contextually relevant content³.

Generative AI can be thought of as an advanced spellchecker, enhancing not just spelling and grammar but also the overall style and syntax of our writing. This technology can assist authors in producing more polished and professional manuscripts, ensuring that their work adheres to high standards of clarity and readability⁴. For instance, it can identify overly complex sentences or awkward phrasing and suggest improvements, making the text more comprehensible to a broader audience⁵. AI can play a crucial role in demystifying our jargon, which is frequently cited as a barrier to understanding for general readers or those working outside the field in question. By identifying and offering simpler alternatives to

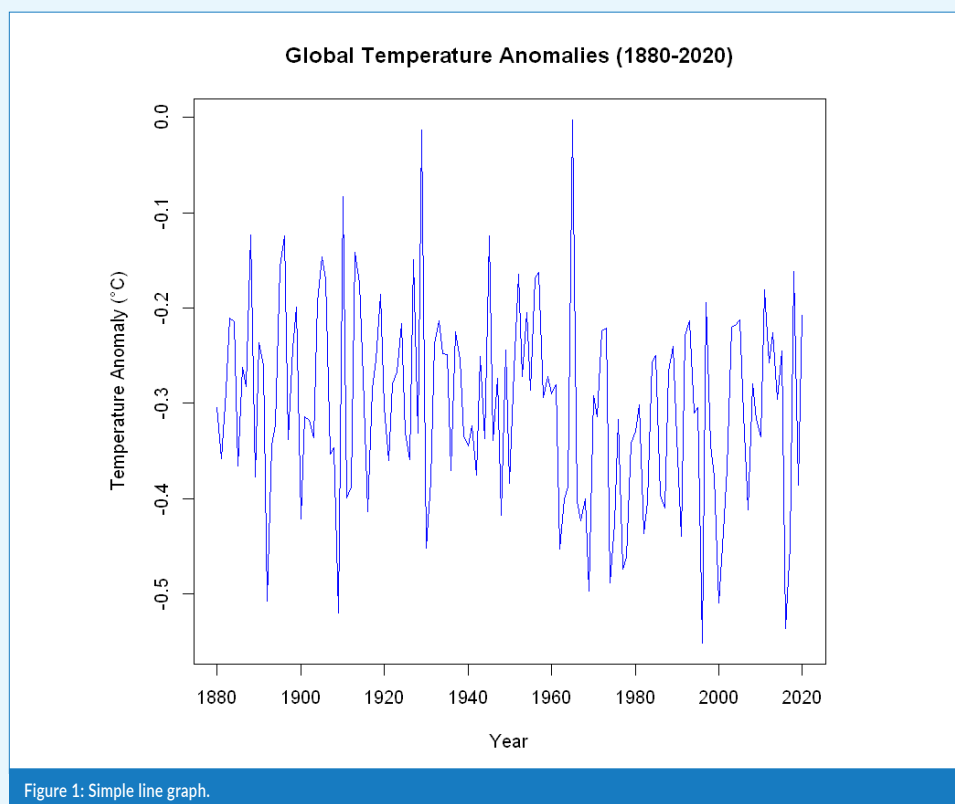


Figure 1: Simple line graph.

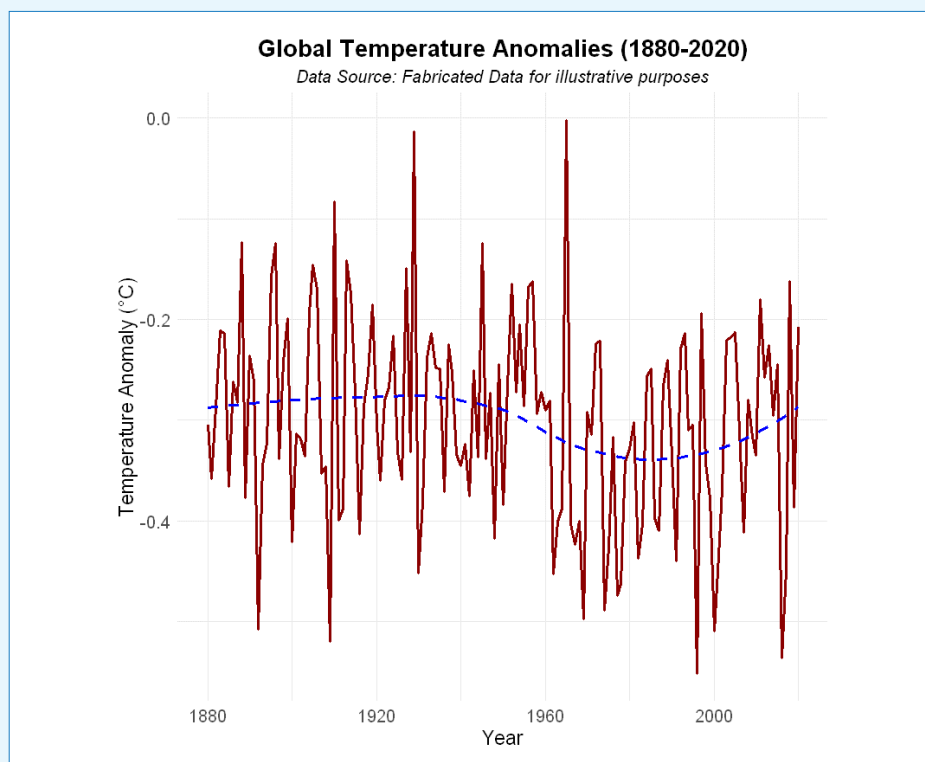


Figure 2: Graph produced with AI assistance to optimise aesthetics and intelligibility.

technical terms (which we readily overlook being so frequently exposed to them) AI helps to ensure that medical literature is more inclusive and accessible. This is in keeping with the goals of the Plain English Campaign, which advocates for clear and straightforward communication to make information more available to the public⁶.

AI is not limited to human speech either. It has demonstrated a remarkable ability to understand and work with programming languages. Anyone who has spent several frustrated hours searching for an elusive misplaced bracket in 'R'⁷ or debugging complex code will appreciate the value of AI-driven tools that simplify these tasks, achieving in a few seconds what could take hours or even days of work. AI systems can not only identify syntax errors and suggest corrections but also optimise code, making it more efficient, easier to read, and less likely to fail when repurposed. It can offer critique of study design, and suggest the most appropriate statistical tests to meaningfully analyse data. It can also offer ideas on how to improve the aesthetics of graphically displayed data. For example, Figure 1 is a simple line graph created using R by simply entering X and Y-axis values from a list. Figure 2 is the result of copying the code into ChatGPT and asking it to make it more visually appealing. In addition to altering the colour palette to a more aesthetically pleasing choice than that produced by the basic statistical software

programme, it has automatically included a LOESS (Locally Estimated Scatterplot Smoothing) line, and has added a trend line that helps the reader to visualise the overall pattern in the data, smoothing out short-term fluctuations to highlight longer-term trends. By automating routine coding tasks, AI allows researchers and clinicians to focus on more complex and creative aspects of their work, enhancing productivity and reducing the risk of human error.

However, the use of AI in manuscript writing and study design is not without its challenges. One significant concern is the potential for AI-related bias. AI models are trained on large datasets that may contain inherent prejudices present in the source material. These can inadvertently be reflected in the AI-generated content, potentially perpetuating stereotypes, skewed perspectives, or factual inaccuracies⁸. The onus will remain on authors to identify these and exclude them from their work. One approach is for developers to use diverse and representative training data, ensuring that the AI is exposed to a wide range of perspectives and experiences⁹. Ultimately, our oversight is crucial; reviewers must consider all submissions in the light of the possibility that it is at least in part AI-generated content, in order to identify and highlight any biased or inappropriate material¹⁰.

While AI models have made significant strides in generating coherent and contextually relevant text, they are not infallible¹¹. AIs are

people pleasers, and would rather generate a sentence riddled with factual inaccuracy than one which doesn't answer the question put to it. There is always a risk of inaccuracies or errors in the content they produce. To mitigate this risk, it is vital to implement robust validation and verification processes¹². Researchers should crosscheck AI-generated content against reliable sources and ensure that any information included is accurate and evidence-based. This 'due diligence' is especially crucial in medical literature, where inaccuracies can have serious implications for patient care and treatment outcomes.

Some critics argue that using AI to generate text is a form of laziness, suggesting that it allows researchers to circumvent the hard work of writing¹³. This however overlooks the complexities involved in effectively utilising AI tools. Writing appropriate AI prompts and rigorously checking the generated content is inherently labour-intensive. Crafting precise prompts requires a deep understanding of the subject matter and clear communication skills. Reviewing AI-generated text demands careful scrutiny to ensure accuracy, coherence, and the elimination of biases. Thus, the process of using AI in manuscript preparation is far from lazy; it requires significant intellectual engagement and effort.

Beyond formal manuscripts and study designs, AI also has the potential to support the reflective practice of surgeons through personal journal entries. Reflective practice is a well-recognised component of continuing professional development in medicine¹⁴. It involves critical self-examination of one's experiences and practices in an effort to improve future performance and patient care. However, events such as the Bawa-Gaba case, have highlighted the perceived (or actual) risks associated with appearing to admit fault in electronic portfolios¹⁵. Resident doctors, in particular, now feel reluctant to document their reflections honestly for fear of legal repercussions or professional censure¹⁶.

In this context, AI can serve as a valuable tool for trainees in that it can assist in rewording journal entries to help them express their reflections without leaving them with the feeling of having incriminated themselves. By providing suggestions for phrasing and terminology, AI (with appropriate prompting) can help surgeons maintain a balance between honesty and caution, ensuring that their reflective practice remains a valuable tool for professional growth without exposing them to undue risk. Surgeons are as heterogeneous a group of people as any other, and as such not all surgeons are naturally gifted >>

writers. For some, particularly those with dyslexia, the task of writing reflective entries can be disproportionately time-consuming and challenging. Some who excel in other areas of their practice might find themselves spending excessive time crafting and revising reflections. AI can alleviate this burden by generating the bulk of the text, which can then be manually edited to align with their personal experiences and technical details of the case in question. This approach allows surgeons to engage more efficiently in the reflective process without detracting from its value, ultimately enabling them to allocate more time to other critical aspects of their professional development.

One concern regarding AI-generated reflections is whether they are genuinely meaningful. Critics argue that AI-generated reflections allow a trainee to appear to have reflected on a subject when, in fact, they have not engaged with the reflective process themselves. However, even if a trainee has not directly written the reflection, they are still exposed to an alternative perspective on the matter, which is inherently valuable. This exposure can prompt further contemplation and learning. While the perspective provided by AI is not human, it offers a different lens through which to view their experiences and challenges, potentially leading to new insights and professional growth¹⁷.

Turnitin, the most widely used plagiarism detection software in British universities, has recently adapted its technology to address the rise of AI-generated content. As AI is increasingly used to generate academic work, Turnitin has developed capabilities to detect content that may have been created or heavily assisted by AI by trying to identify the specific patterns and linguistic features that are characteristic of AI-generated text^{18,19}.

The company has introduced new algorithms and machine learning models designed to flag AI-generated content, which typically exhibits a different structure, vocabulary, and syntax compared to human-written text. However, this is a rapidly evolving field, and both AI generation and detection technologies are continually being refined. As AI tools become more sophisticated, they will almost certainly adapt to outmanoeuvre comparatively crude plagiarism detection software, even if it is itself AI-driven. At present though, unrefined AI output can be easily detected by Turnitin and similar products²⁰.

Is that all that impressive?
Let's prompt ChatGPT 4.0 to summarise the last two paragraphs:

'This ongoing 'arms race' between AI content generation and detection highlights the dynamic nature of academic integrity in the digital age. Universities and educators must

stay informed about these technologies to effectively manage and mitigate the risks of AI-assisted academic misconduct.'

There is no doubt that the above makes grammatical sense; the AI has done its job. Would you be fooled into thinking a human wrote it though? The overly formal tone raises the first red flag, but in an academic context, it might not be inappropriate. The clichéd phrasing is a bigger give-away. Beyond the context of an overly enthusiastic marketing pitch, who would describe academic integrity as 'dynamic'? The content (or rather, the conspicuous lack of content) is another clue. Unless it is carefully fed specifics, AI-generated text often stays within the realm of generalisations to avoid errors, making it seem vague and non-committal. The output speaks in broad terms without delving into specific details or examples that might be expected in human-written text. This, combined with the general sentence structure (they follow a predictable pattern that mirrors the training data of the AI), all just feels a little too robotic. It adheres closely to learned structures without the kind of creative or unconventional approaches a human writer might take.

This is an example of the Uncanny Valley effect, a concept originating from robotics and artificial intelligence. The Uncanny Valley refers to the unsettling feeling people experience when something non-human, such as a robot or AI-generated content, closely resembles human behaviour or appearance but still has subtle differences that make it seem off-putting or eerie. The term was first introduced by Japanese roboticist Masahiro Mori in 1970²¹. He observed that as robots became more human-like, people's emotional response to them became increasingly positive, until a certain point – where the robot was

almost human but not quite – where people felt unease or discomfort instead of empathy (see Figure 3). This dip in emotional response is what Mori termed the 'Uncanny Valley'. Some have speculated that natural selection preserved this response in our ancestors as it kept them away from the recently deceased (contagion-avoidance), or even other (potentially hostile) hominid species²². In the context of AI-generated text, the Uncanny Valley effect manifests when the content is nearly indistinguishable from human writing but still contains cues that reveal its artificial origins, which we cannot help but subconsciously detect.

There is also the question of whether it is even necessary to cite AI or even mention it as a contributor to the work. To return to the analogy of AI as simply a sophisticated spellchecker; just as researchers are not expected to cite the use of the Microsoft Office spellcheck or thesaurus functions when they are utilised, it is not necessary to specifically cite AI tools when they are used to refine and polish a manuscript. The primary focus should remain on the content and the research itself, with AI serving as a background tool to support the writing and planning processes. Unedited or unrefined AI-generated output results in the sort of robotic and non-committal prose seen in the above example.

The integration of generative AI in medical literature, study methodology design, and reflective practice offers a wide variety of benefits, but it is essential to remain cognizant of its flaws. By implementing robust oversight and validation processes, and fostering a supportive environment for AI's integration into reflective practice, we can harness the full potential of AI while mitigating its risks.

As AI technology continues to evolve, its role in medical research and practice is certain to expand, offering new opportunities for innovation and improvement across many fields, including our own.

In coming years, it is highly likely that learning to create efficient AI prompts will be as vital a skill for the budding researcher as is learning to properly reference a source or creating an appropriate graph to explain their data.

Surgeons who fail to adapt to these emerging tools risk being left behind. ■

References

References can be found online at www.boa.ac.uk/publications/JTO.



Figure 3: AI-generated image of a traditional Japanese Bunraku puppet. Its eerily 'nearly human' features may provoke the revulsion response known as the Uncanny Valley phenomenon.