

AI for medical writers: Friend or foe?

Jamie Norman¹, Lisa Chamberlain James²

¹ TriloDocs Ltd., Cambridge, UK

² Trilogy Writing & Consulting Ltd., Cambridge, UK

doi: 10.56012/qovt1204

Correspondence to:

Jamie Norman

Jamie.norman@trilogywritng.com

Abstract

Artificial intelligence (AI) is beginning to affect almost every industry, and medical writing is no different. But how does this relate to our industry? How will AI affect medical writers? What's already available and what is in the pipeline? Should medical writers be happy and embrace the technology, or should we resist as much as we can, assuming that we will all be replaced by machines? This article discusses the current state of the art of AI in medical writing and asks the question: AI for medical writers – friend or foe?

How did we get here?

What a year it's been for artificial intelligence (AI) already! The pace at which the conversation around AI has accelerated in just a few short months is unprecedented. However, AI is certainly not new. As a term, AI was coined back in the 1950s,¹ and ever since then, the technology, models, and processing power have advanced. With ChatGPT leading the way, along with Google, Meta, and a host of other tech companies, the paradigm is shifting so rapidly that in the time between writing this article and publishing it, there could be something new to discuss in the world of AI.

But what led us to this point? What triggered this explosion? AI is not new nor are language models such as those employed by ChatGPT. As we enter the age of AI, and with ChatGPT competing with the behemoth of Google, the success is best explained by Google's own history.

In the early days of the internet, conducting a "search" seemed like something of a dark art. Companies would invest their marketing budgets in promoting their URL because the idea

of just being able to search for the company seemed to be a pipe dream. Even with the advent of the first search engines, if you did not know how to write queries using Boolean logic, getting any meaningful results felt like a lottery.

And then Google came along: no pop-up ads, no confusing page layout, just a simple search box. And it worked. Effortlessly. The beauty was in how they made something so complex incredibly simple and accessible. And the rest, as they say, is search history. And now history repeats itself: AI is not new, but a simple, well-designed-interface such as ChatGPT makes it appear effortless and provides powerful results. This has captured the imagination of the world. It is certainly impressive and has

prompted a flood of examples demonstrating its power. As Arthur C. Clarke famously said, "Any sufficiently advanced technology is indistinguishable from magic".²

What was once a niche domain for data scientists and AI technologists has suddenly become widely accessible.

What was once a niche domain for data scientists and AI technologists has suddenly become widely accessible. We now see everyone leveraging its power for everything from drafting emails to answering exam questions. This explosion has been so large and rapid that it has outpaced working practices and even legislation. This has led to the kind of concerns that triggered the open letter from tech leaders in which they urged a pause in development of AI to allow some checks and regulations to be put in place.³



What AI is and how it works in a writing context

In a rapidly changing sector, what is already available and for what purpose?

The term AI is very broad. Different branches of it often get conflated, but there are disciplines within the discipline. At its highest level, AI is a catch-all term for any computational technique that enables machines to mimic human behaviour. This could be as simple as a macro in excel that automatically performs a set of calculations or procedures or as advanced as a facial recognition algorithm.

The next layer of detail is referred to as “machine learning”, which is a subset of AI that uses statistical methods to improve a model based on experience. For example, for image recognition, this could be a system that improves the accuracy of recognising a certain animal under increasingly ambiguous scenarios.

The next deeper level is so-called “deep learning”. It is a subset within machine learning, where a neural network is used to make connections. Incredibly large, multi-layered

networks create computational systems that work more like the human brain. Many deep learning algorithms are actually closer to “black box systems”, in which the outcomes may be incredibly accurate but difficult to explain. This is one of the areas that makes some groups pause because they often show emergent behaviours that were not predicted by humans and can be unsettling, adding to concerns that AI is out of control.

This is where the notion of “explainable AI” comes in.⁴ Being able to reverse-engineer outcomes and explain the results of AI models creates a more comforting outcome, although this may mean sacrificing some of the computational power provided by deep learning models.

Where does ChatGPT fit in?

ChatGPT uses neural nets to support the computation power of its outcomes. As a large language model, it retains a degree of “explainability”.⁵ Large language models generally use statistical models. In simple terms, a language model uses a set of training data to create a

probability of the next word or series of words in a sentence. ChatGPT’s power comes from access to perhaps the largest corpus of training data of any language mode. However, even ChatGPT has shown emergent behaviours. For example, it can be used to solve maths problems, for which it was not specifically designed, and although it can “solve” maths problems, it cannot interpret statistics.

Language modelling also cannot assign probabilities to linguistically valid sequences that may not have been in the training data. This is a positive in the sense that it can create novel texts, but it also can produce results that are grammatically correct but factually incorrect. That is, it can assess the probability of word sequences but cannot understand their meaning. In this way, language models differ from cognitive models, which, as their name suggests, are closer to our own abilities to solve problems.

The challenge of interpreting new concepts is an important consideration for AI. This has been illustrated using the “Monty Hall” problem from the medium of game shows.⁶ The Monty Hall problem is a brain teaser, in the form of a probability puzzle, loosely based on the American television game show “Let’s Make a Deal” and named after its original host, Monty Hall.

Imagine that you are given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say number 1, and the host, who knows what’s behind the doors, opens another door, say number 3, which reveals a goat. He then says to you, “Do you want to pick door number 2?” Is it to your advantage to switch your choice?

Most people’s intuition is to stick with their original choice. However, the correct response is, counter-intuitively, to switch. Switching gives a two in three probability of winning a car, while sticking with your original choice gives only a one in three chance. If you do not believe it, there are plenty of referenceable articles on this topic that can be found on Google.

If you pose this question to ChatGPT, you will receive the correct response, suggesting that you switch. This is due to the training data, which most likely included a written reference to how this problem is solved. However, what if we made this a “dumb” problem, where the answer is much more obvious? In this case, we pose the same problem but with a small change: This time the

The challenge of interpreting new concepts is an important consideration for AI.



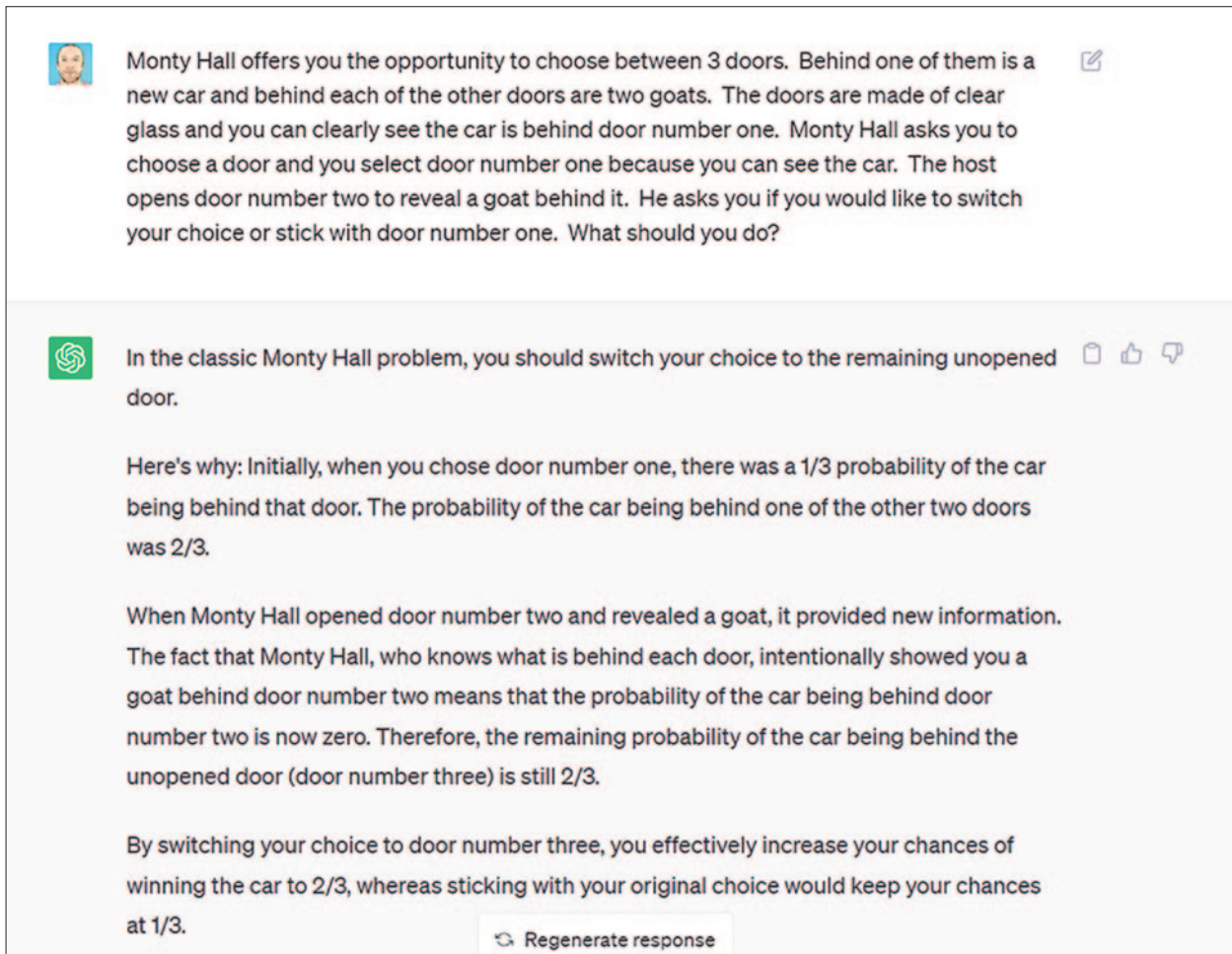


Figure 1. GPT prompt and response to the “dumb” Monty Hall Problem

With thanks to Colin Fraser, Data Scientist at Meta.

doors are made of clear glass so that you can see behind every door. Under these conditions, you can easily pick the door with the car behind because you can see it, and when asked to switch this time, you would clearly stick with your choice. However, when posing this challenge to ChatGPT, it always suggests switching your choice (Figure 1).

This is a reflection of the language model’s inability to reason in the same way as a human – to make deductions from premises or to process insights rather than to make probabilistic inferences from word frequencies. This explains why making new inferences from data can be challenging, and it is exactly the kind of challenge we face in interpreting statistical data from new drugs. The margins for error in this context are significantly smaller so we cannot rely on

The challenge of interpreting new concepts is an important consideration for AI.

language models alone.

Like any technology, ChatGPT is just a tool. As with any tool, it is only as good as the person using it. ChatGPT is incredibly powerful, but to build products around it, its underlying working models, nuances, and other details need to be understood.

How could AI help medical writers?

Many generic language models are able to create authentic content, but they do not always perform well when the content is novel or its frame of reference is new, as was the case with the dumb Monty Hall problem previously mentioned.⁶ This is simply a result of the training data used because language models can only produce content related to the data they have

been trained on. A well-documented downside of generic language models is “computer hallucinations”, where a language model “makes up” information or cites references when it has no information. This is obviously a major concern for the field of scientific writing.

To address this, some niche tools have been specifically trained on and produce content relating to scientific information. An example is Perma ai,⁷ which searches the abstracts of papers to answer specific text-based questions and can support research scientists. Another is BioGPT,⁸ which is a spin off from ChatGPT designed specifically for life sciences and produces more relevant biological text. Our own tool, TriloDocs,⁹ combines a sector-specific language model with a core of expert rules to provide a set of “guidelines” and only interprets relevant information from clinical trial data in relation to specific best practice criteria.

It seems that the future of AI in the medical writing sphere may not be as stand-alone tools but rather within platforms that use it in the context of wider rules and other elements. Using AI tools in the medical writing space as more of a “walled garden” makes sense because of reluctance to upload intellectual property, personal data, or other sensitive information to open platforms, where data ownership and data protection are currently being debated. Regulatory authorities need to be confident in the accountability and traceability of raw data and documents supporting any claims. GDPR (General Data Protection Regulation), protection of commercially sensitive information, and “AI hallucinations”, not to mention the specific context of medical writing, remain major concerns.

Nonetheless, language models are undoubtedly powerful tools for creating authentic-looking texts from certain prompts, rewriting texts for different audiences (e.g., in other languages), and producing simplified summaries. Most medical writers would be delighted to pass on routine, mundane, and repetitive tasks to a computer, which can do them more efficiently, accurately, and quickly. This could liberate writers to concentrate on the highly skilled tasks of contextualising and interpreting clinical data and allow them to have meaningful data discussions with clinical teams much earlier than is currently possible. In the med-comms and medical journalism worlds, AI tools can help writers more quickly and accurately create time-sensitive documents and sift through huge amounts of literature.

What are the risks of AI?

We have already touched on some of the key risks involved in using AI. **Data privacy** is often the main risk that springs to mind. However, this is an inherent risk of any technology and not specific to AI. Some AI platforms present a risk of being internet-based. Also, “open” systems present a risk even in a non-AI context. Some emerging options allow developers to build a language model within a secure environment

(although the training data are publicly available). How this develops in the medical writing arena will be interesting.

Risk of errors. In our experience with TriloDocs, the risk of human error has been significantly reduced, if not eliminated. Important data that humans may miss are identified by the tool, and we have not yet found an issue raised during quality assurance that was not already identified by the technology. The problem of AI hallucination is a cause for real concern because there is no room for false data, inferences, or references when dealing with clinical and scientific data. The more niche platforms will have to specifically eliminate this risk, which may pose a significant challenge.

From a medical writing perspective, a conservative approach is always best. Our experience is that it is better for the tool to highlight where something is missing or interpretations cannot be made, flagging data points for the medical writer to investigate rather than having a tool that produces a “complete” but misleading draft.

Other considerations include the **ethical debate** about AI, which is far outside the scope of this article. Jamie Bartlett,¹⁰ a journalist and author specialising in technology and a regular speaker on the topic of futurism, has warned that only three things can be guaranteed about the future of technology: firstly, that data storage capabilities and demand will continue to grow at an exponential scale; secondly, that the processing power of computing will also continue to grow, which along with the ability to store huge amounts of data, has powered this latest AI revolution; and thirdly and most importantly, that human drives and behaviours will not change.

The limiting factor to AI is how we implement these tools and how ethically we can introduce checks and balances to manage them. There is almost an AI paradox playing out

in front of us: We all want AI to help us to do our jobs better or at least take away the more menial parts of our work without replacing us altogether. Unfortunately for some, that choice will not be theirs to make.

What does all this mean for medical writers?

One thing we always stress when talking about our own platform, TriloDocs, is that it does not replace the medical writer. TriloDocs simply accelerates and enhances the writer’s ability to have meaningful data discussions with the clinical team and speeds crafting of the report. We have not yet met anyone who actually enjoys trawling through data with a highlighter pen and interrogating tables for information; crafting a strong narrative around the data, however, is an entirely different proposition.

Highly skilled medical writers bring value as critical thinkers as they create study reports and related documentation. We are still some way off from the ultimate goal of AGI (Artificial General Intelligence), which moves AI into the realm of human-like thought. Until that point, critical thinking can only be done by humans. In the short time that tools like Chat CPT have captured our imagination, there is already an adage that describes where things could be going in the short term: AI might not take your job, but someone who uses AI will.¹¹

AI is not going away – medical writers cannot influence that – but we can influence how we approach and use AI. If we view AI as a tool that can supplement our work, make us more efficient and accurate, and relieve us of some of the heavy lifting, then it can become a powerful resource, freeing us to focus on the more valuable work of critical thinking and crafting a strong narrative in our highly complex and vital work.

Acknowledgements

The authors gratefully acknowledge Dr Barry Drees’ input on this article.

Like any technology, ChatGPT is just a tool. As with any tool, it is only as good as the person using it.

A well-documented downside of generic language models is “computer hallucinations”, where a language model “makes up” information or cites references when it has no information.

The limiting factor to AI is how we implement these tools and how ethically we can introduce checks and balances to manage them.



References

1. McCarthy J, Minsky ML, Rochester N, et al. A proposal for the Dartmouth summer research project on artificial intelligence. 1955 [cited 2023 Jul 10]. Available from: <https://web.archive.org/web/20070826230310/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.
2. Clarke AC. Profiles of the future: An inquiry into the limits of the possible. 1st ed. New York: Harper & Row; 1973.
3. Future of Life Institute. Pause giant AI experiments: an open letter. 2023 [cited 2023 Jul 10]. Available from: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.
4. Google. Explainable AI in industry (tutorial). 2023 [cited 2023 Jul 11]. Available from: <https://sites.google.com/view/explainable-ai-tutorial>.
5. Potts C. Stanford Webinar: GPT-3 & beyond. 2023 [cited 2023 Jul 10]. Available from: <https://www.youtube.com/watch?v=-lnHHWRCDGk>.
6. Fraser C. ChatGPT: Automatic expensive BS at scale. Medium. 2023 [cited 2023 Jul 10]. Available from: <https://medium.com/@colin.fraser/chatgpt-automatic-expensive-bs-at-scale-a113692b13d5>.
7. Ferma. The quickest path to your next eureka. 2023 [cited 2023 Jul 10]. Available from: <https://www.ferma.ai/>.
8. Microsoft/BioGPT. BioGPT: Requirements and installation. 2023 [cited 2023 Jul 10]. Available from: <https://github.com/microsoft/BioGPT>.
9. Trilogy Writing & Consulting Ltd. TriloDocs AI-enhanced medical writing: The AI tool for clinical study reports. 2023 [cited 2023 Jul 10]. Available from: <https://trilogywritng.com/trilodocs>.
10. Bartlett J. The People vs Tech: How the internet is killing democracy. 1st ed. India: Ebury Press; 2018.
11. Confino P, Burton A. AI might not replace you, but a person who uses A.I. could. Fortune. 2023 Apr 25 [cited 2023 Jul 10]. Available from: <https://fortune.com/2023/04/25/artificial-intelligence-ai-replace-humans-prompt-engineers-chatgpt/>.

Author information

Jamie Norman is the Chief Product Officer of TriloDocs GmbH. Jamie began his career in marketing, working with several life sciences brands. In 2010, he made the transition to technical product management and has since worked in product leadership roles for several high-growth and enterprise software brands specialising in data science and AI-led products.



Lisa Chamberlain James is a CEO of Trilogy Writing & Consulting Ltd. Following a PhD and postdoctorate at the University of Cambridge, in 2000, she began a medical writing career. She has since been involved in EMWA as a member of the Educational Committee, mentor, leader, and assessor of workshops, and she teaches and reviews workshops for the American Medical Writers Association. Lisa holds an EMWA Professional Development Certificate, is a visiting lecturer for Kings College, London, initiated and chaired the EMWA Pharmacovigilance and Communicating with the Public Special Interest Groups, is Chair of EMWA's Geoff Hall Scholarship Committee, Section Editor for *Medical Writing*, and a Fellow of the Royal Society of Medicine.