

Scientific writing at the dawn of AI

Jarosław Meyer-Szary¹, Miłosz Jaguszewski², Szymon Mikulski³

¹Department of Pediatric Cardiology and Congenital Heart Diseases, Faculty of Medicine, Medical University of Gdańsk, Gdańsk, Poland

²First Department of Cardiology, Medical University of Gdansk, ul. Smoluchowskiego, Gdańsk, Poland

³Department of Head & Neck Surgery, Singapore General Hospital and National Cancer Center Singapore; Duke-National University of Singapore (NUS) Medical School, Singapore

The following article serves as a companion piece to “Current Understanding of Duchenne Muscular Dystrophy. A Purported Interview with a Purported Expert,” published herein [1], which in fact was written by an Artificial Intelligence (AI) chatbot as an experiment and, to our knowledge at the time of writing, is the first article published in a peer-reviewed medical journal to be fully generated by AI.

Experiment

The purpose of our experiment was to test whether the latest publicly available AI-enabled natural language processing (NLP) tools are capable of authoring a medical article. While the short answer is not yet, AI is indeed able to compose notes on medical topics for inexperienced readers, or to assist scientists in writing full-fledged scientific papers. The following commentary explains the methodology of our experiment and offers our insights into the implications of AI on the future of scientific writing.

As the AI platform, we opted for Open AI's Playground and its latest most capable model text-davinci-003 (GPT-3.5 series), a state-of-the-art large language model (LLM) available for public use as an online web application (<https://platform.openai.com/playground>), wherein the user asks questions by typing and submitting prompts, and AI responds in text within 3–4 seconds (sic), similar to an online live chat.

The initial aim of our experiment was to have AI perform and write up a full-length literature

review. However, after initial testing, this expectation was proven unrealistic. Overall, the AI's output featured acceptable language, had the standard structure and style, but lacked the expected depth of detail and factual expertise.

When prompted to provide a list of citations, the AI returned links to topically accurate web pages. When prompted for the references in the Vancouver citation style, the output seemed appropriate in format at first glance, but upon scrutiny transpired to be nonsense in content. Although the journal names were mostly correct, the issue and page numbers were wrong. Similarly, the referenced authors were real but frequently never published together or in the relevant field. Finally, the cited article titles were all fictitious despite seemingly being on topic. Examples of such citations are available in the Appendix, Output #2. Thus, the AI-provided citations were found completely unreliable and were removed.

Next, the authors prompted the AI chatbot to provide a summary of a particular scientific article, the complete main text of which (excluding the abstract, figures, and tables) was submitted as a prompt. The resultant output closely resembled a typical abstract both in its length and content. Most of the information within was correct, but some generalizations or misinterpretations led to factually incorrect statements (see Appendix, Output #4). Those wishing to engage AI in the generation of abstracts for their own papers must appreciate this important shortfall and perform a thorough fact-check of the output. Hence, while

Address for correspondence: Jarosław Meyer-Szary, Department of Pediatric Cardiology and Congenital Heart Diseases, Faculty of Medicine, Medical University of Gdansk, Skłodowskiej-Curie 3a, 80-210, Poland, e-mail: jmeyerszary@gumed.edu.pl, tel: +48 583492882

Received: 21.02.2023

Accepted: 31.05.2024

This article is available in open access under Creative Commons Attribution-Non-Commercial-No Derivatives 4.0 International (CC BY-NC-ND 4.0) license, allowing to download articles and share them with others as long as they credit the authors and the publisher, but without permission to change them in any way or use them commercially.

Table 1. The strengths and weaknesses associated with using AI in scientific writing.

Advantages	Disadvantages
Speed and efficiency	Lack of originality
Can generate drafts quickly	May produce content that lacks new insights
Reduces time spent on routine tasks	
Accessibility of knowledge	Quality and accuracy issues — hallucinations
Helps synthesize vast amounts of data	May require extensive fact-checking
Makes complex information more accessible	Risk of propagating errors
Enhanced productivity	Dependence on data quality
Supports researchers in brainstorming and drafting	Outputs are only as good as the input data
Allows more time for critical thinking and analysis	
Language assistance	Potential for bias
Provides language and writing aid for non-native speakers	May inadvertently perpetuate existing biases
Improves clarity and coherence of writing	
Innovative perspectives	Ethical and authorship concerns
Can suggest novel connections between ideas	Raises questions about authorship and intellectual property
Encourages interdisciplinary approaches	Ethical considerations in attribution

it may be of assistance, the current AI technology remains inadequate in stand-alone summarizing and interpretation of scientific literature.

Taking the above limitations into consideration, the authors proceeded to further constrain the task given to AI. Rather than a systematic review of the literature or an abstract of a full-text scientific article, the AI chatbot was next tasked to generate a “transcript” of a short interview with a content expert on the topic of Duchenne muscular dystrophy (DMD). The resultant output was deemed suitable for attempt at publication.

To arrive at the final text, multiple prompts were attempted in a trial-and-error process until the output satisfied the authors in terms of linguistic soundness and factual correctness. Nonetheless, some of the generated answers were incorrect (e.g., claiming that the most common arrhythmia found in DMD is bradycardia, whereas in fact it is various forms of tachycardia), while others were incomplete (e.g., omitting the crucial role of respiratory support in the discussion of DMD treatment).

Hence, the final interview article was compiled from AI responses to the following 3 prompts:

1. Write 10 questions regarding Duchenne Muscular Dystrophy for an interview with a professional medical expert. Answer those questions in formal, scientific language, each answer should be about 400–500 words. Cite sources.

2. Write a scientific paper on Duchenne Muscular Dystrophy focusing on treatment advances using formal, scientific language. Cite sources in the Vancouver citation style.
3. What are the respiratory management strategies including the role of nocturnal assisted breathing and mechanical ventilation in Duchenne Muscular Dystrophy?

The authors’ original input was limited to conceptualization, composition of the prompts, fact-checking of output against available literature and domain knowledge, and minimal editing (question and answer numbers were added for legibility). References were removed for the reasons explained above. The 3 prompts above were submitted sequentially to obtain a satisfactory composite response. The initial output to Prompt 1 was inadequate in its discussion of treatment of DMD, hence Prompt 2 and Prompt 3 were added, respectively, to instruct the AI to expound on that topic and then specifically on respiratory management of DMD, which was initially completely omitted from Output 2.

Commentary

The progression of artificial intelligence (AI) from the realm of science fiction to reality has accelerated in recent years and culminated in AI becoming widely and freely available to lay internet users throughout the world. In their latest

versions, AI agents driven by natural language processing models such as OpenAI's GPT-3 are capable of generating text responses indistinguishable at first glance from content thought out and composed by an intelligent human. In doing so, modern AI is finally reaching for the holy grail of computer science pursued since the birth of modern computing.

Perhaps the first to famously consider the implications of AI was Alan Turing (1912–1954), a renowned English mathematician, computer scientist, and philosopher. In his 1950 paper “Computing Machinery and Intelligence,” [2] the polymath described his eponymous Turing test as an imitation game: a test of a machine's ability to exhibit intelligent responses indistinguishable from those of a human.

Today, ChatGPT, the blockbuster AI agent fine-tuned for conversations using relatively short text answers, which was launched for public use in November 2022, has arguably come the closest to achieving just that — passing the Turing test. In at least one earlier study of ChatGPT, the best judgment of human participants was no better than a coin toss in correctly attributing short news pieces to a human or AI author [3]. These results have astounded not only computer scientists and creative writers but have also piqued the interest of popular global media, with multiple dedicated articles bringing the ongoing developments in AI to the lay masses.

Interestingly, relatively little attention has been paid to this phenomenon by the medical scientific community, especially from the perspective of scientific writing and publishing. Given AI's recent meteoric rise in capability, availability, and prominence, it is safe to assume that AI-driven writing tools are becoming the new reality, and it is unrealistic to expect that the AI genie will return to the bottle. Hence, as a follow-up to our companion piece published in this issue “Current Understanding of Duchenne Muscular Dystrophy. A Purported Interview with a Purported Expert,” herein the authors discuss a few implications of the recent developments in AI, which may be relevant specifically to the medical scientific community.

Advantages

No doubt, the upsides of using AI in scientific writing are multiple and difficult to miss. Scientists will welcome the time savings afforded by the boost in efficiency conferred by AI

over tedious tasks such as sifting through and summarizing countless sources. This in turn is expected to lead to improved quality of scientific output by enabling scientists to focus their expertise on higher-level cognitive tasks such as data interpretation, study design, or knowledge application.

Although our experiment has demonstrated that the current technology is still limited in its ability to effectively analyze long and complex text and to generate meaningful, error-free insights, it is a matter of time before automated research will aid scientists in quickly and accurately researching topics across vast bodies of knowledge spanning multiple scientific disciplines. We hope that in doing so, AI-assisted research will become more expeditious and free from availability bias.

Aside from content generation, the strength of AI collaboration lies in its tirelessness and unparalleled processing speed, which can be harnessed to provide rapid review and real-time editorial feedback to authors to identify errors, minimize inconsistencies, and improve accuracy.

Moreover, although creativity itself was once expected to long remain an exclusive domain of human writers, it is quickly becoming a part of AI's repertoire of skills. Scientists will undoubtedly look to their AI collaborators to help generate new ideas, ask unexpected questions, and inspire novel ways of thinking.

Finally, the current medical scientific process is a manpower- and resource-intensive, and costly endeavor. At the same time, ongoing advances in computing power, its non-stop global availability, and ever-improving capability confer AI with a potential multiplier effect that far surpasses that of any human team, and, initial capital investments aside, at a fraction of the cost of conventional methods.

Overall, AI-enhanced research and scientific writing will bring advantages not only to individual authors and the scientific community, but ultimately also to patients, who should in turn benefit from timelier and more affordable access to a larger interdisciplinary knowledge base and scientific progress in the form of AI-enabled breakthroughs and discoveries.

Current limitations

But let us not get ahead of ourselves. The future of medical scientific writing with AI also poses novel challenges and confers certain downsides.

First and foremost, current AI engines are primarily language models and do not perform fact-checking against accepted scientific sources or databases. As a result, though at first glance accurate-sounding and probable, AI-generated content, however sophisticated, may not be justified by its training data and is prone to factual inaccuracy. This deceptive convincingness is a common AI phenomenon referred to as hallucination.

Second, the skeptics are not wrong to quickly judge AI-generated work for its lack of creativity and nuance compared to human writing. After all, current AI technology is inherently derivative, “merely” using predetermined mathematical and statistical models to repeat the most probable quanta of information included in its training dataset, reformatting, and relaying them as its own output. And for that, AI has been nicknamed a “stochastic parrot”.

Furthermore, although AI models have the capacity to evolve and learn, their “ability” is constrained and biased by the training dataset – both in its scope and accuracy as a source.

The current platforms including GTP-3 are trained on vast open-source internet repositories, and not directly on scientific databases of peer-reviewed publications. Hence, any semblance of scientific authority that AI models may mimic are based on derivative works of variable accuracy found throughout the internet. Specifically, the training dataset for GTP-3 consists of 60% web archive, 22% Reddit archive, 16% books archive, and 3% Wikipedia [3]. Notably, the training occurs at a fixed point in time (June 2021 in the case of GTP-3.5 series), beyond which the AI “knowledge-base” is frozen in time and rendered potentially outdated.

Solutions to the issue of stochastic mimicry and hallucination are at the forefront of current AI advances. As of this writing, next-generation tools are becoming available, which couple the latest natural language processing models with AI-driven web and database real-time search engines (Microsoft’s Bing and Alphabet’s Bard). In doing so, AI technology has arguably leapt past Turing’s imitation game and continues to inch ever closer to perhaps the final holy grail of artificial general intelligence (AGI), whereby a computer will no longer be constrained to a particular domain such as language and numerical processing. Instead, AGI will be capable of learning and performing any intellectual task that a human being ever could, and likely beyond.

Future implications

Technological limitations aside, and whether we like it or not, the era of AI content generation has unequivocally arrived. It therefore behooves us to contend with its implications. As with any disruptive technology, the arrival of AI in the scientific community is poised to shake up its established order. It remains to be seen whether AI will act as a unifier, helping to level the playing field, or whether it will further discriminate between resource-poor and -rich settings.

On the one hand, the latest AI technology demands immense computational power both to train the models and then to resolve the queries, which in turn requires significant electrical energy and capital investment. Consequently, those with more abundant resources are likely to have an edge over scientists from smaller institutions and those from less developed countries. Likewise, industry-sponsored projects may potentially take the lead over those driven by academia.

On the other hand, many AI resources are made freely available to anyone on the internet in open or beta versions – for now. However, AI developers will undoubtedly want to capitalize on their investment, and access to the most powerful implementations is likely to become limited to prosperous users in the form of paid services.

Thus, as the power of AI becomes increasingly marketable, one can expect the next generation of AI tools to come at a premium, the cost of which will create another source of inequality in the global scientific community.

Another open debate pertaining to the implications of AI in scientific writing is on the moral-ethical consequences of its use. In other words, will so-called AI ghostwriting be perceived as downright academic dishonesty or misconduct? Or does the question merely boil down to appropriate acknowledgment and disclosure?

Although not strictly illegal, undisclosed ghostwriting in general is widely criticized by the scientific community. At the same time, publishers (e.g., British Medical Journal) may consider ghostwriting acceptable if properly acknowledged, and the European Medical Writers Association has published guidelines on the use of professional medical writers to ensure ethical and responsible practice [4]. It is their stance that ghostwriters’ expertise is effective in conveying scientific data, which in turn leads to improved quality papers.

Should similar principles be applied, and analogous ghostwriting license extended to AI

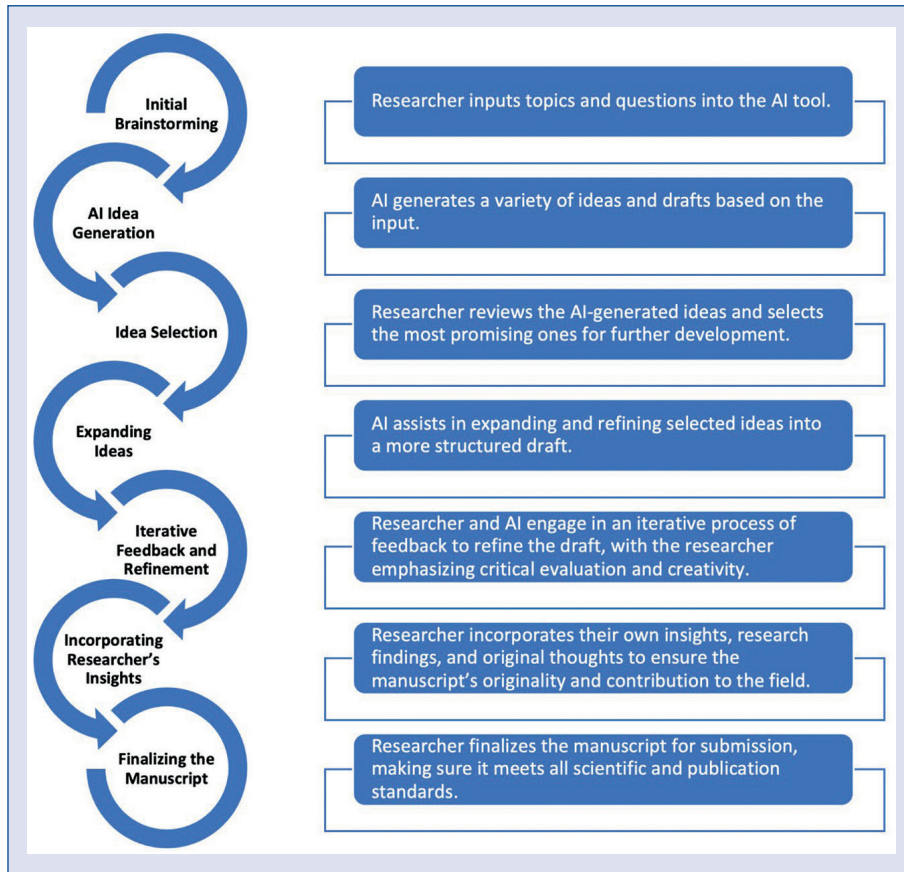


Figure 1. Flowchart illustrating the process of AI-assisted scientific writing.

co-authors? It is our belief that given the rapid rise in AI capability and ubiquity, AI assistance in scientific writing is a foregone conclusion. Hence, regulators will need to decide not if but to what extent AI authorship will be permissible and how to ensure honest disclosure and appropriate attribution.

In the meantime, editors themselves may, paradoxically, turn to technology as a means of detection of AI-generated content through automated linguistic analysis of submitted text passages. One such tool is AI Detector from the company Writer [5]. Having analyzed our companion piece “Purported Interview with a Purported Expert,” the software flagged it as “71% human-generated,” whereas in fact it is nearly 100% AI. Fortunately, the current commentary article received a passing score of “100% human.”

Conclusions

With the blockbuster application ChatGPT reportedly attaining a record-breaking 100 million active users in under two months, the exponential growth in the capability and availability of AI-driven

natural language processing tools is poised to revolutionize all disciplines in which writing plays a significant role, medical scientific writing included. Whether we extol its apparent virtues or lament the folly of surrendering to AI, we must remember to retain an open yet simultaneously skeptical mind to the opportunities ahead. Although predictably some consequences of the dawn of AI will be welcomed and others will not, many remain yet unforeseen.

References

1. Meyer-Szary J, Mikulski S. Current understanding of Duchenne muscular dystrophy — a purported interview with a purported expert. *Cardiol J.* 2024; 31(3): 367–368, doi: [10.5603/cj.94330](https://doi.org/10.5603/cj.94330).
2. Turing AM. Computing machinery and intelligence. *Mind.* 1950; LIX(236): 433–460, doi: [10.1093/mind/lix.236.433](https://doi.org/10.1093/mind/lix.236.433).
3. Brown TB, Mann B, Ryder N, et al. Language Models are Few-Shot Learners. *ArXiv abs.* 2020, doi: [10.48550/arXiv.2005.14165](https://doi.org/10.48550/arXiv.2005.14165).
4. Jacobs A, Wager E. European Medical Writers Association (EMWA) guidelines on the role of medical writers in developing peer-reviewed publications. *Curr Med Res Opin.* 2005; 21(2): 317–322, doi: [10.1185/030079905X25578](https://doi.org/10.1185/030079905X25578), indexed in Pubmed: [15802003](https://pubmed.ncbi.nlm.nih.gov/15802003/).
5. <https://writer.com/ai-content-detector/> (13.02.2023).