Artificial or Augmented Authorship? A Conversation with a Chatbot on Base of Thumb Arthritis

Ishith Seth, BBiomed(Hons), MD†‡§
Peter Sinkjær Kenney, MD, PhD¶
Gabriella Bulloch, BMSc+, David J. Hunter-Smith, MBBS(Hons), MPH, FRACS, FACS*‡§
Jørn Bo Thomsen, MD, PhD*‡¶
Warren M. Rozen, MBBS, BMedSc, MD, PhD, FRACS*‡§

Summary: ChatGPT is an open artificial intelligence chat box that could revolutionize academia and augment research writing. This study had an open conversation with ChatGPT and invited the platform to evaluate this article through a series of five questions on base of thumb arthritis to test if its contributions and contents merely add artificial unusable input or help us augment the quality of the article. The information ChatGPT-3 provided was accurate, albeit surface-level, and lacks analytical ability to dissect for important limitations about base of thumb arthritis, which would not be conducive to potentiating creative ideas and solutions in plastic surgery. ChatGPT failed to provide relevant references and even “created” references instead of indicating its inability to perform the task. This highlights that as an AI-generator for medical publishing text, ChatGPT-3 should be used cautiously. (Plast Reconstr Surg Glob Open 2023; 11:e4999; doi: 10.1097/GOX.0000000000004999; Published online 26 May 2023.)

INTRODUCTION
Chat Generative Pre-Trained Transformer 3 (GPT-3) (OpenAI, Calif.) is an open artificial intelligence (AI) platform that utilizes deep learning to produce human-like text, which could greatly reduce the time spent on literature search, data analysis, and research writing in the future, while ensuring academic standards of writing.1,2

AI-supported writing tools using machine learning algorithms have been prevalent since 2005, with SCIgen being the earliest to use this technology for generating figures and research writing for MIT.3 Although instances of AI-generated research articles that have been peer-reviewed and accepted for publication exist, they generally involve a human researcher collaborating with an AI system, rather than the AI working autonomously.2,4 The lack of peer review and common flaws upon providing information has largely alienated AI platforms; however, in January of 2023, ChatGPT made its research debut, listed as a lead author in the Nurse Education in Practice journal.2 This has raised ethical questions for scientific communities about the inseparability of human and ChatGPT-3 derived abstracts, especially considering they went undetected through plagiarism softwares.5,6 In addition, the ChatGPT-3 training database is limited until 2021, which prevents it from incorporating research evolving on a daily to monthly basis. The use of large language models, such as ChatGPT, can potentially augment researchers’ work in scoping and developing their research area and scientific articles. In this exponentially growing field, little is still known as to the value these models provide for scientific work. We therefore evaluated ChatGPT’s responses in this article to assess its potential utility in academia.

METHODS
This study prompted a conversation with ChatGPT on a popular plastic surgical topic, base of thumb arthritis, to evaluate the specificity of its generated answers. A total of five questions were posed to ChatGPT, and each subsequent query was predicated on the response provided.
to ascertain whether its answers were improved. The aim of this study was to evaluate ChatGPT’s use for research writing in the plastic surgery field. Its input was reviewed by experts (D.H.S, W.R, and P.K), and scrutinized for how updated, accurate, and succinctly presented the information was in text and referencing.

**RESULTS**

The first prompt was, “In 200 words, describe the current evidence on surgical management for base of thumb arthritis with relevant references” (Fig. 1). ChatGPT provided a relevant and accurate summary of different surgical management of base of thumb arthritis with minimal grammatical or structural errors, and only 1% was detected in Turnitin as plagiarism. No references were provided with the prompt.

The following prompt was, “In 200 words, describe the current evidence for implants in the surgical management of base of thumb arthritis, describing the quality of evidence, and provide five references” (Fig. 2).

ChatGPT provided a more specific answer when prompted about current evidence for implants for base of thumb arthroplasty and provided relevant level I evidence. The authors noted that ChatGPT would only provide references when prompted a specific number such that

**Takeaways**

**Question:** Is ChatGPT (an open artificial intelligence platform) useful enough to contribute to scientific research or does it merely add artificial unusable input?  
**Findings:** ChatGPT-3 provided a quick and relevant summary of the subject with minimal grammatical or structural errors but failed to provide relevant references and even “created” references instead of indicating its inability to perform the task.  
**Meaning:** Overall, it was clear the information ChatGPT-3 provided was accurate, albeit surface-level to contribute to scientific research.
In 200 words, describe the current evidence for implants in the surgical management of base of thumb arthritis, describing the quality of evidence and provide five references.

Base of thumb arthritis is a common condition that affects the joint at the base of the thumb, causing pain and difficulty with grasping and gripping objects. Surgical management of this condition includes the use of joint implants, which can be used to replace or repair the affected joint.

The current evidence for the use of implants in the surgical management of base of thumb arthritis is generally considered to be high quality. Studies have shown that joint implants can effectively reduce pain and improve function in patients with this condition. For example, a randomized controlled trial published in the Journal of Hand Surgery (Am) in 2016 found that patients who received a joint implant experienced significant improvements in pain and function compared to those who received a non-surgical treatment.

Other studies have also demonstrated the safety and effectiveness of joint implants for the treatment of base of thumb arthritis. A systematic review published in the Journal of Hand Surgery (European Volume) in 2016 found that joint implants had a high rate of success and low complication rate.

In conclusion, the current evidence for the use of joint implants in the surgical management of base of thumb arthritis is generally considered to be high quality, with multiple studies demonstrating their safety and effectiveness.

References:
3. Surgical management of trapeziometacarpal osteoarthritis: a systematic review [Journal of Hand Surgery (European Volume), 2016.]

Fig. 2. Prompt: In 200 words, describe the current evidence for implants in the surgical management of base of thumb arthritis, describing the quality of evidence, and provide five references.
In 200 words, describe the current evidence for implants in the surgical management of base of thumb arthritis, describing the quality of evidence using Centre for Evidence-Based Medicine levels of evidence, and using 5 references of level 4 Centre for Evidence-Based Medicine evidence or higher and offer an evidence based approach to management.

Base of thumb arthritis is a common condition that causes pain and difficulty with gripping and grasping activities. Surgical management of this condition can include the use of implants, such as trapeziometacarpal (TMC) joint implants.

The current evidence for the use of TMC joint implants in the surgical management of base of thumb arthritis is primarily based on level 4 evidence from case series and case-control studies. These studies have demonstrated that TMC joint implants can be an effective treatment option for reducing pain and improving function in patients with base of thumb arthritis. However, the quality of evidence is limited by the small sample sizes and short-term follow-up in these studies.

One high quality level 4 study by Beaudreuil et al. (2015) reported on a series of 37 patients who underwent TMC joint arthroplasty for the treatment of trapeziometacarpal osteoarthritis with a mean follow-up of 4.6 years. They found that the procedure resulted in a significant improvement in pain and function scores, with 89% of patients reporting satisfaction with the procedure.

Another high quality level 4 study by Badois et al. (2016) reported on a series of 24 patients who underwent TMC joint arthroplasty for the treatment of trapeziometacarpal osteoarthritis with a mean follow-up of 4.4 years. They found that the procedure resulted in a significant improvement in pain and function scores, with 96% of patients reporting satisfaction with the procedure.

In summary, the current evidence for the use of TMC joint implants in the surgical management of base of thumb arthritis is primarily based on level 4 evidence from case series and case-control studies. These studies have demonstrated that TMC joint implants can be an effective treatment option for reducing pain and improving function in patients with base of thumb arthritis. However, the quality of evidence is limited by the small sample sizes and short-term follow-up in these studies. Therefore, a multidisciplinary approach that includes physiotherapy, occupational therapy, non-surgical interventions and patient education is recommended before considering surgery.

In this answer, ChatGPT stated multiple avenues for base of thumb arthritis management that could be focused on in future research. ChatGPT also stated that biological therapies had the potential to slow the progression of arthritis and improve joint functions; however, examples like platelet-rich plasma, stem cells, and hyaluronic acid investigated in phase 2 and 3 clinical trials have shown no difference in comparison with placebo or corticosteroids, and were limited to small sample size and short follow-up. Therefore, ChatGPT provided no innovative ideas for advancing the management of base of thumb arthritis and advocated for known therapies which had no real evidence for their use.

The last prompt was “In 200 words, where is the lack of evidence in management of base of thumb arthritis;
provide relevant references” (Fig. 4B). Here, ChatGPT raised valid points on the paucity of evidence in management and provided superficial ideas on research that is needed to fill in the gaps, for example, the timing of surgery, the optimization of surgical implant, and long-term follow-up data. Despite this, no ideas were ones that had not been previously discussed.

**DISCUSSION**

In this article we evaluated ChatGPT-3’s ability to base of thumb arthritis. On the surface, ChatGPT-3 provided a quick and relevant summary of the subject with minimal grammatical or structural errors but failed to provide relevant references and even “created” references instead of indicating its inability to perform the task. This highlights that as an AI generator for medical publishing text, ChatGPT-3 should be used cautiously. The references about current evidence in literature provided by ChatGPT-3 seem to be rather random, as important high level evidence references are left out. This is of cause a major flaw when performing a literature search, and makes its input rather useless for research writing. It is more dangerous to receive information that is deceiving than to receive information that is outrageously incorrect without subtlety. Researchers who are willing to take the risk of its input at face value could promote obviously incorrect or even made up research. OpenAI is well aware of this issue, and improvement of this would result in upvalue in academia.

Furthermore, it is clear that ChatGPT-3 cannot think “outside the box,” when prompted about providing recommendations for future treatment, highlighting that the current platform is devoid of innovation or creativity. It is worthy to highlight that creativity is an aspect of human information processing which is not well understood neuroscientifically and is poorly produced by AI algorithms so far. In addition, the lack of new information available to the platform probably discounts the novelty of its ideas, and perhaps if ChatGPT had access to the most up-to-date resources its ideas would be considered more innovative. This supports its use as a supplementary resource for writing and refining, rather than a first author that is expected to produce ideas and present it in a novel and interesting way in context of other emerging information. Even so, considering how far this technology has come already, it is possible the next versions could augment and innovate in ways we cannot yet understand.

**CONCLUSIONS**

Overall, it was clear the information ChatGPT-3 provided was accurate, albeit surface-level. It seems that while the algorithm has access to all information on the internet until 2021, it lacked analytical ability to dissect for important limitations about base of thumb arthritis, which would not be conducive to potentiating creative ideas and solutions in plastic surgery. Therefore, it is the authors’ opinions that ChatGPT-3 could promote convergent thinking and prevent innovation. For this purpose, in its current version
ChatGPT-3 should be limited within the scope of research, or at least reviewed under the guidance of experts.

Peter Sinkjær Kenney, MD, PhD
Department of Plastic Surgery
Odense University Hospital
Odense 5000
Denmark
E-mail: p@rhinx.com

DISCLOSURE
The authors have no financial interest to declare in relation to the content of this article.

ACKNOWLEDGMENT
Chat-GPT3 was used to generate some of the content in this work.

REFERENCES
5. Lopez T, Qamber M. The benefits and drawbacks of implementing chatbots in higher education: a case study for international students at Jönköping University. 2022. Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:chj:diva-57482