



A deeper dive into ChatGPT: history, use and future perspectives for orthopaedic research

Matthieu Ollivier, Ayoosh Pareek, Jari Dahmen, M. Enes Kayaalp, Philipp Winkler, Michael Hirschmann, Jon Karlsson

► To cite this version:

Matthieu Ollivier, Ayoosh Pareek, Jari Dahmen, M. Enes Kayaalp, Philipp Winkler, et al.. A deeper dive into ChatGPT: history, use and future perspectives for orthopaedic research. *Knee Surgery, Sports Traumatology, Arthroscopy*, 2023, 31 (4), pp.1190-1192. 10.1007/s00167-023-07372-5 . hal-04480903

HAL Id: hal-04480903

<https://hal.science/hal-04480903>

Submitted on 27 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A deeper dive into ChatGPT: history, use and future perspectives for orthopaedic research

Matthieu Ollivier¹ · Ayoosh Pareek² · Jari Dahmen^{3,4,5} · M. Enes Kayaalp⁶ · Philipp W. Winkler⁷ · Michael T. Hirschmann⁸ · Jon Karlsson⁹

The rate of plagiarism and false content in scientific literature varies depending on the field of study and the methods used to detect it. According to some studies, the overall rate of plagiarism in scientific literature is estimated to be around 2–3% [1]. However, the rate can be higher in certain fields and for certain types of content. Additionally, the rate of false or fraudulent content in scientific literature is difficult to quantify, as it often goes undetected or is not reported. However, cases of scientific misconduct, including the fabrication and falsification of data, have been reported in various fields and can have serious consequences for both the authors and the scientific community.

It is important for the scientific community to maintain high standards of ethics and accuracy in scientific research to ensure the validity and reliability of the published literature.

In this editorial, we will touch upon the description of Large Language Models, define their limits and strengths and finally explore options to detect fraudulent manuscripts.

Large Language Models (LLM) have the potential to assist researchers in generating clear and concise writing, summarising vast amounts of information and performing various language-related tasks [2]. This can potentially save

researchers' time and improve the efficiency of the scientific writing process. However, it is important to note that the output of language models like ChatGPT should always be critically evaluated for accuracy and scientific validity, as they are not capable of independent scientific reasoning or experimentation. Ultimately, the impact of language models on scientific writing will depend on how they are adopted and utilised within the scientific community*.

The GPT (Generative Pretrained Transformer) was first introduced by OpenAI in 2018 as a language model that uses deep learning techniques to generate human-like text. It was trained on a large corpus of text data, allowing it to generate coherent and contextually relevant responses to various prompts [3].

GPT-1, the first version of the model, was soon followed in 2019 by GPT-2, which was even larger and more powerful, with the ability to generate entire articles and perform various language tasks such as translation and summarisation. GPT-3, the latest version, was released in 2020 and it has set new standards in the field of AI language generation, achieving state-of-the-art results in various benchmark tests and becoming the largest publicly available language model

✉ Matthieu Ollivier
ollivier.mt@gmail.com; ollivier.matthieu@yahoo.fr

¹ Aix Marseille Univ, CNRS, ISM, Institute Movement Science, Marseille, France

² Sports Medicine and Shoulder Service, Department of Orthopedic Surgery, Hospital for Special Surgery, 535 East 70Th Street, New York, NY 10021, USA

³ Amsterdam UMC, Department of Orthopaedic Surgery, Amsterdam Movement Sciences, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands

⁴ Academic Center for Evidence Based Sports Medicine (ACES), Amsterdam, The Netherlands

⁵ Amsterdam Collaboration for Health and Safety in Sports (ACHSS), International Olympic Committee (IOC) Research Center Amsterdam UMC, Amsterdam, The Netherlands

⁶ Department for Orthopaedics and Traumatology, Istanbul Kartal Dr. Lutfi Kirdar Training and Research Hospital, Istanbul, Turkey

⁷ Department of Orthopaedics and Traumatology, Kepler University Hospital Linz, Linz, Austria

⁸ Department of Orthopedic Surgery and Traumatology, Head Knee Surgery and DKF Head of Research, Kantonsspital Baselland, 4101 Bruderholz, Bottmingen, Switzerland

⁹ Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska University Hospital, Sahlgrenska Academy, Gothenburg University, Gothenburg, Sweden

at the time. It has been widely adopted in various industries and applications, including chatbots, virtual assistants and content generation. As a highly advanced language model, ChatGPT has several strengths that make it useful in scientific project writing. First, its ability to generate contextually relevant responses allows it to assist in the creation of clear and coherent writing. Additionally, it has been trained on a vast corpus of text data, including scientific content, which enables it to have a good understanding of scientific terminology and writing conventions. Moreover, ChatGPT is able to generate large amounts of text quickly, making it an efficient tool for summarising scientific findings or creating drafts**.

The system has its limits in scientific project writing. As it is based on statistical patterns in its training data, it may produce biased or inaccurate information if the data it was trained on contain biases of this kind. Additionally, whilst it is able to generate scientifically plausible text, it is important critically to evaluate and verify its output, as it may not always be scientifically accurate. Lastly, ChatGPT is not capable of independent scientific reasoning or experimentation and is only able to generate text based on patterns it has seen during its training*.

One example is if you prompt ChatGPT with the task of writing a scientific paper on a new hypothesis about the causes of a certain disease. Although ChatGPT may be able to generate scientifically plausible text based on its training data, it will not be able to reason independently about the validity of the hypothesis or design experiments to test it. This is because ChatGPT is only able to generate text based on patterns it has seen in its training data and does not have the capability to engage in independent scientific thinking or experimentation. It is up to the human user critically to evaluate the generated text and determine its scientific validity*.

More Large Language Models like ChatGPT are able to hallucinate and make up information and appear confident doing so.

Johann Mikulicz-Radecki was a Polish-Austrian physician and researcher who made important contributions to surgery, particularly in the field of orthopaedic surgery. Born in 1850 in Austrian-ruled Poland, he studied medicine in Vienna and later became a professor at the university.

Mikulicz-Radecki was known for his surgical innovations and techniques, particularly in the treatment of musculoskeletal disorders. He was one of the first surgeons to perform joint replacement surgeries and was a pioneer in the use of metal implants for this purpose. He also made important contributions to the surgical treatment of spinal injuries and deformities, developing new techniques for spinal fusion and instrumentation.

Finally, to provide exceptional performance, language models like GPT-3 are trained on large amounts of data. GPT-3, for instance, has been trained on 45 terabytes of

text data from multiple sources. However, the training data are usually from a specific time frame and may not reflect recent events or developments*.

As the author and editor of orthopaedic scientific journals, we were thrilled to evaluate the potential of ChatGPT. The single question of allowing “it” as a co-author for scientific publication is a fiercely debated subject [4, 5].

We, therefore, need to anticipate the high risk associated with the extensive use of language models to improve (or even create from the scratch) orthopaedic publications [6].

There are several options to enhance our ability to verify the validity and reliability of scientific content. Some of these include the following.

1. Data sharing: encouraging the sharing of raw data and methods used in scientific research may increase transparency and allow for the greater independent verification of results.
2. Improved training and education: providing researchers with training in research ethics, scientific integrity and data management can help to ensure that they adhere to high standards and produce reliable and valid research.
3. Improved technology and tools: developing new technology and tools to detect plagiarism, fraud and other forms of scientific misconduct can help improve the accuracy and reliability of scientific content.

Several AIs are already available to analyse fragments of text and estimate the potential use of Large Language Models to improve or even create scientific content.

The creators of ChatGPT themselves have designed a specific tool [7] designed to facilitate discussions about the difference between human-written and AI-generated content. The results of the model should be taken into consideration but not relied upon solely as proof of whether a document was created by AI. The model was trained using text written by humans from multiple sources, which may not cover all the forms of human-written text. Each document is classified as very unlikely, unlikely, unclear, possibly, or likely to be generated by AI.

When asking AI Text Classifier to analyse the fragment on the life of Johann Mikulicz, it classified the document as unclear.

More recently, blockchain technology has been proposed to enhance the security and originality of scientific projects [8, 9]. Here are a few ways in which blockchain can be used to achieve this.

- Traceability of scientific work: blockchain can be used to create an immutable record of the origin, development and evolution of scientific projects, making it easier to track their progress and verify their authenticity.

- Management of intellectual property: blockchain can be used securely to manage and track the intellectual property associated with scientific projects, ensuring that original ideas and findings are properly credited and protected.
- Encryption of sensitive data: blockchain can be used securely to store and manage sensitive scientific data, ensuring that the data are only accessible to authorised individuals and that their confidentiality is protected.
- Detection of plagiarism and misconduct: blockchain can be used to detect plagiarism and other forms of scientific misconduct by allowing for the secure and transparent tracking of scientific information.

By utilising blockchain technology, the scientific community is able to enhance the security and originality of scientific projects and ensure that the published literature is reliable and trustworthy. This can help to build trust and confidence in the scientific process and ensure that the results of scientific research can be relied upon**.

Throughout this manuscript, parts marked * were written entirely by ChatGPT and not edited by the authors, whilst parts marked ** were partially written by ChatGPT and edited by the authors.

This technology can be seen as a modern Pandora's box for scientific writing and the box seems to be already open. Barriers are needed to avoid the contamination of the orthopaedic scientific community and detect scientific misconduct. AI-doped blockchain systems used extensively to avoid fraud in various domains might be the future of scientific editing.

Authors' contributions MO, AP, JD, MEK, PWW, MH, JK, MH drafted the manuscript and completed the necessary revisions. All the authors have reviewed and edited the manuscript.

Funding Not applicable.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval Not applicable.

Informed consent Not applicable.

References

1. Weber-Wulff D (2019) Plagiarism detectors are a crutch, and a problem. *Nature* 7749:435. <https://doi.org/10.1038/d41586-019-00893-5>
2. Gordijn B, Have HT (2023) ChatGPT: evolution or revolution? *Med Health Care Philos.* <https://doi.org/10.1007/s11019-023-10136-0>
3. Cahan P, Treutlein B (2023) A conversation with ChatGPT on the role of computational systems biology in stem cell research. *Stem Cell Rep.* <https://doi.org/10.1016/j.stemcr.2022.12.009>
4. Graham F (2023) Daily briefing: ChatGPT listed as author on research papers. *Nature.* <https://doi.org/10.1038/d41586-023-00188-w>
5. Thorp HH (2023) ChatGPT is fun, but not an author. *Science.* <https://doi.org/10.1126/science.adg7879>
6. Else H (2023) Abstracts written by ChatGPT fool scientists. *Nature.* <https://doi.org/10.1038/d41586-023-00056-7>
7. AI TEXT CLASSIFIER. OpenAI API. <https://platform.openai.com/ai-text-classifier>
8. Boetto E, Golinelli D, Carullo G, Fantini MP (2020) Frauds in scientific research and how to possibly overcome them. *J Med Ethics medethics* 6:2020–106639. <https://doi.org/10.1136/medethics-2020-106639>
9. Kumar R, Singh D, Srinivasan K, Hu YC (2022) AI-powered blockchain technology for public health: a contemporary review, open challenges, and future research directions. *Healthc Basel Switz* 11(1):81. <https://doi.org/10.3390/healthcare11010081>