RMD Open
Rheumatic & Musculoskeletal Diseases

VIEWPOINT

The wide range of opportunities for large language models such as ChatGPT in rheumatology

Thomas Hügle

To cite: Hügle T. The wide range of opportunities for large language models such as ChatGPT in rheumatology. *RMD Open* 2023;**9**:e003105. doi:10.1136/rmdopen-2023-003105

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/rmdopen-2023-003105).

Received 24 February 2023
Accepted 14 April 2023

Check for updates

© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Department of Rheumatology, University Hospital Lausanne (CHUV), University of Lausanne, Lausanne, Switzerland

Correspondence to
Professor Thomas Hügle; thomas.hugle@chuv.ch

Now that's what you call technical disruption: When has a technical invention ever spread so quickly since the introduction of the iPhone in 2007? One million users in 5 days… It's like the world has been eagerly waiting for this one application. OpenAI's ChatGPT has hit the widest scale and is being celebrated by the press. ChatGPT explains, analyses and creates knowledge from the internet, replacing Google's pure keyword search in no time. And it is reduced to the mother of all user interfaces: the chatbot. ChatGPT creates creative texts as well as fact-based scientific essays and can automatically programme other algorithms. Anyone can use it, from the average consumer to the quantum researcher… Gone are the days when we almost bashfully searched Google instead of PubMed for medical evidence or case reports. ChatGPT is now also integrated in Microsoft's search engine Bing and combines keyword search with syntax and the possibility of interactive conversations. Is our need for convenience so great that, with speech recognition almost mature, we could also use automated speech generation? Or is ChatGPT (I'd like to say &Co, but there is no &Co yet) really this incredible lever on all levels of society, including medicine and research? And why did the artificial intelligence (AI) architecture of the 'generative pretrained transformer' (GPT) so easily outgun previous algorithms such as Google's Bidirectional Encoder Representations from Transformers and triggered a code red there?

Back to the beginning. It is Sunday 11 December 2022 in Paris at the Expo Porte de Versailles. I am in the queue to register for the French Congress of Rheumatology. A friend of mine (a psychiatrist) calls me excitedly because he knows I am scientifically interested in AI and we are working together on a project about digital treatment for fibromyalgia and depression. He tells me about ChatGPT: 'This chatbot changes everything, it is simply brilliant linguistically, even in German'!

Over 3 months have passed since then. And it's true: The machine creates unlimited, real-time fact-based texts in a way that no human could ever do. From poems to introductions or discussions of scientific texts. ChatGPT can programme in Python and create algorithms, which could perhaps help to make better use of clinical data, for example.

HOW DOES THE ALGORITHM WORK?

ChatGPT is a version of the language model GPT3 (a 'natural language processing' algorithm) that has been specifically trained for improved 'chat' capabilities. Meanwhile, there has already been an update to an even more performant version GPT4. The fine details behind how it is trained and used have not been made public, but this is what we know so far: the GPT architecture is, as the name suggests, a pretrained AI model that uses a neural network (here, a transformer) to generate new text based on previous texts. The pretraining is done using an 'unsupervised' method. This involves feeding a huge data set of text from books, newspaper articles, blogs, social media sites such as Reddit, etc and asking GPT to predict the next word in the text.

In order for language models such as GPT to understand words (which have little inherent meaning for a computer), they must convert the words into vectors of numerical values, called word embeddings, which typically include 300 and 1500 different values for each word. These values represent different 'traits' of a word—a simplistic example being the 'youngness' or a word vs the 'oldness' of the word. Representing words as values like this enables models like GPT to understand

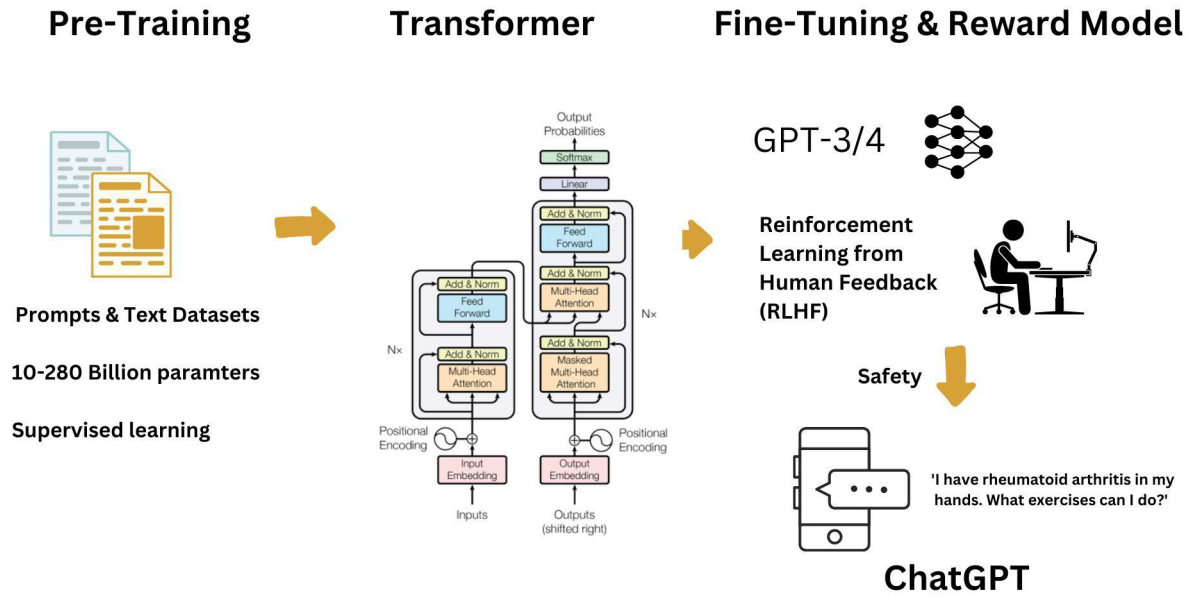

Figure 1 ChatGPT algorithm. Adapted from OpenAI. The ChatGPT is a hybrid algorithm in which a model is first generated from a large text dataset through supervised learning. Then, with the help of human labellers, a reward model is attached that makes the text output even more similar to human speech.

which words are similar even if they use different letters (eg, 'rapid' and 'speedy'). An 'encoder' network is used to convert words into values, and a decoder converts these vectors back into meaningful text (figure 1).

To progress from GPT3 to ChatGPT, a technique called 'reinforcement learning with human feedback' was used. Reinforcement learning involves providing a 'reward' based on desired performance and was also used in chess computers such as Deepmind's AlphaZero. For ChatGPT, this involved an intermediate human step, where different model outputs were ranked by humans in terms of quality. This intermediate human step is important, however, and serves to check the output text for comprehensibility and meaningfulness. Finally, security precautions are taken to ensure that ChatGPT does not produce any potentially harmful content.

HOW I USE CHATGPT AS A RHEUMATOLOGIST AT THE MOMENT

I use ChatGPT3 since December 2022. ChatGPT4 has been launched while editing this article. In the beginning, I have used it every day, for text formatting (typos, shortening, summarising, etc) searching for references (attention, confabulations occur, see below), helping to write introductions (not for this one), creation of non-scientific medical texts and creation of Excel or Python codes (rather gimmicky so far).

If ChatGPT3 were an employee, my job reference after 2 months would look like this: 'ChatGPT is a proficient text generator. He/she/it assembles information securely as texts in a wide variety of languages. Spelling and grammar are largely error-free. The language engine tries to recognise connections, but does not always succeed. ChatGPT usually recognises its limitations, for example, when recommending therapies, and refers to medical professionals. The summarisation of texts is reliable. In a literature search, ChatGPT unfortunately so far only records references up to the year 2021, which is insufficient from a scientific point of view. In some cases, references created by ChatGPT cannot be identified in PubMed. ChatGPT is extremely hard-working, can be used 24/7 and is never sick. However, for the last month or so, there have been recurring symptoms of overwork, especially during the day, but this will stop in the future through payment (so far ChatGPT works for free). ChatGPT can apparently programme its own codes and algorithms, but we don't quite understand that yet as clinicians. Our trust in ChatGPT is growing, but we can't yet have confidential data analysed by them or simply copied into the user interface. We need to talk to its parent OpenAI first and instal the code in our clinic to keep the data secure. ChatGPT is well liked by its staff for its skills and diligence, but it has no sense of humour.

WHAT IS THE RESEARCH ON CHATGPT IN HEALTHCARE?

So far, there is no clear evidence yet on how precisely ChatGPT may advance medicine. To date (March 2023), there are 36 medical publications on PubMed about the topic; 5 of them in 2022 and already 31 in 2023. Most of them are position papers and discussions from the domain of education, medical writing or automation. Many of these publications have been critical. In a recent publication, the journal Science, takes a flippant view: 'ChatGPT is fun, but not an author'.[1] Nature takes a more critical view and states in its Daily briefing: 'Science urgently needs a plan for ChatGPT'.[2] A controversy has quickly opened up in titles such as 'ChatGPT-friend or foe?', 'Abstracts written by ChatGPT fool scientist', or 'Chatbots are a double-edged sword'.[3–6] Other

publications see chatbots as a great danger to science or even its 'greatest enemy'.[7] Nature takes a step forward and demands that certain basic rules be observed when dealing with chatGPT.[3] In contrast, no published work exists yet on how chatGPT specifically helps with medical treatment. There are still no reports that clinical study protocols created by ChatGPT are successful, that ChatGPT brings quality to medical reports or it helps with clinical decisions. In my view, the first area in which scientific evidence of ChatGPT will be generated is the area of therapeutic education and health literacy.

A STRONG SYNERGY: CHATGPT AND ELECTRONIC MEDICAL RECORDS

In contrast to research and education, ChatGPT is seen much more positively in automation. For example, at Epic, a market leader in electronic medical records (EMR). As an EMR, Epic collects data from hundreds of millions of patients. Although only partially structured and in a lot of free text, more medical data comes together here than in any medical register in the world. It is no wonder that Epic, with its 'Cosmos' or 'Better Care' programmes, is training algorithms to learn through AI from the many millions of clinical decisions made and make them useful for individual patients. ChatGPT, as a champion of processing and generating free text, comes at just the right time. But before ChatGPT may turn into a clinical decision-support system, it will help on another level. The magic formula is called 'Automation Workflow Systems'. Combining ChatGPT and EMR may become a weapon against the rampant bureaucracy. Once the algorithm has the necessary data from the EMR available (medical history, laboratory, X-ray findings, etc) consultation reports, discharge reports but also cost-credits for health insurance companies, certificates of incapacity to work, etc could all be created in real time and only need to be validated. The relief for medical and administrative staff would be enormous. For doctors and nurses alone, several hours of work per week on documentation would be eliminated, which can be much better used for patient consultations or further training.

CHATGPT HAS FUELLED THE NO-CODING REVOLUTION

AI for solving specific, often particularly repetitive tasks, has spread through all industries, including healthcare. Over 500 algorithms are FDA approved, mostly as classification/diagnostic tools in imaging, including one or two rheumatology indications.[8] In the clinic, we have access to a large amount of data and we know the clinical problems which we would like the algorithms to solve. Preferably with a user interface directly in the EMR. The limiting factor is finding programmers or paying them. Therefore, more and more so-called no-coding platforms are coming into use, in which code is generated (eg, in Python) and a user interface is directly included.[9] An even more flexible solution than such platforms could be to have the code written by ChatGPT with the EMR as user interface. However, it may be risky to do so without any oversight, as ChatGPT may not appreciate the full context of the code and there's a risk of security vulnerabilities. Oversight from clinicians with programming skills, or clinical informaticians will likely to be required for the foreseeable future.

RISKS OF CHATGPT

While this viewpoint focuses more on the positive aspects of large language models, there are some important risks. For me, the biggest risk is that ChatGPT is unfortunately not always right. Sometimes, ChatGPT confidently states false information. OpenAI has been open about this from the beginning, however, and recently, it has been explicitly pointed out again on the chatbot. As mentioned earlier, some of the references provided by ChatGPT during my own literature search could not be found in PubMed or Google and seemed to be confabulated. Academic reference errors with concrete examples are also pointed out in the OpenAI API Community Forum.[10] The chatbot sometimes paraphrases, and since it otherwise communicates in a very qualified manner, it is difficult to deduce these paraphrases from the text. It is different with the keyword search via Google, here one has at least the external aspect of the website or credibility of sources within the framework of affiliation, organisation, peer-reviewed publications, etc. In other words, humans are sometimes too gullible towards humans and probably also chatbots. This means ChatGPT delivers bite-sized but ultimately unvalidated information to the user. Other critical points relate to the AI as a whole, not just the chatbot. Where exactly does the data come from? How transparent is the algorithm? Is there a kind of ranking like Google for the content? And finally, what does ChatGPT actually do with the data? If we feed in our own data, ChatGPT keeps learning and becomes uncatchable? The other open question is whether these models continue to improve as more parameters are added—or whether it is now other techniques (such as reinforcement learning from human feedback, more fine-tuning) that will improve performance.

IS CHATGPT A GAME-CHANGER FOR RHEUMATOLOGY?

In March 2023, the results for a PubMed search for the terms 'ChatGPT AND rheumatology' still were 0 (for oncology 6 and psychiatry 5 results appeared). No scientific publications were found in a Google search for the same terms. Conversely, there are various articles about rheumatology and ChatGPT in social media such as LinkedIn, Twitter and in Podcasts. In a recent Tiktok video, the American rheumatologist, Dr. Clifford Stermer, asked a patient with systemic sclerosis to use ChatGPT to create an insurance letter with scientific references for the coverage of an echocardiogram.[11] While the letter was perfectly written, some of the references were made up by the chatbot. In my opinion, this confirms that for the moment, ChatGPT is a well-trained language model,

but not a scientific model. This function will certainly be improved in future versions of ChatGPT, or in language models which are more specifically trained on scientific text such as Googles SciBERT. Anyways, for simple written automatisation tasks such as insurance letters, ChatGPT will definitely be a game-changer and safe healthcare professionals precious time by taking administrative work away from us by using data from EMRs. The time saved for bureaucratic tasks such as cost credits for biological treatment to insurances or discharge reports, could be very large.

CHATBOTS AS INTERFACE FOR PATIENTS

In rheumatology, there is a high proportion of patients with chronic diseases and a long patient journey, but at the same time, an important shortage of healthcare professionals. The delays for a consultation can be immense and the need for information often cannot be met by general practitioners. Chatbots on websites of hospitals, private practices, etc can simplify communication and automate it at least for certain parts. This also applies to telephone chatbots during times when the telephone is not manned, for example, during breaks. On a more didactic level, chatbots may substantially contribute to health literacy in patients with chronic diseases such as arthritis. Chatbots can educate patients, for example, in the form of weekly interactive interventions on lifestyle, physical exercise, drug adherence, etc. ChatGPT provides reasonable information on questions such as 'What is the best diet when suffering from rheumatoid arthritis?'). Or, 'I have rheumatoid arthritis—what exercises can I do for my swollen joints?' For other therapeutic questions, such as drug therapy, ChatGPT answers with a security answer, but then gives some very general but quite helpful answers: 'As an AI language model, I am not qualified to provide medical advice. However, there are certain exercises that can help alleviate symptoms of rheumatoid arthritis (RA), including swollen joints….' On the other hand, cognitive–behavioural therapy or mindfulness elements might well be provided by the chatbot, for example, for patients with primary or secondary fibromyalgia. The query, can you do a general cognitive behavioural therapy exercise with me?' results in a reasonable answer (online supplemental file 1). Importantly, chatbots can be connected or directly asses patient reported outcomes on disease activity, quality of life or digital biomarkers such as biometric data. By a reward function (as part of reinforcement learning) future chatbot models will thus be able to learn which type of suggested intervention was useful or not. ChatGPT thus inevitably moves in the direction of a DIGA (German term for certified Digital Health Applications which are reimbursed by health insurances), although this would of course require regulatory aspects to be worked on and evidence to be shown. Of course, this never comes close to a human coach, who may have less knowledge than ChatGPT, but more senses, empathy, facial expressions, gestures or similar interests and experiences as patients. In any case, chatbots and human coaches could work together on digital platforms on a broader scale.

Figure 2 An image created with DALL.E (OpenAI) through artificial intelligence. The input prompt was: 'Hand with rheumatoid arthritis touching a screen'. A caption on the top left with meaningless text was removed. As so-called multimodal AI, automatic images or sketches can be created in the future to match the text provided by ChatGPT, for example, for educational purposes. AI, artificial intelligence.

MULTIMODAL AI

A further direction of development is multimodal AI, where multiple types of data are combined, such as audio (speech), image, text and time series data. ChatGPT can, for example, work together with speech recognition tools to create reports with automatically inserted prognoses or lay descriptions for patients. Or it could merge with DALL.E, the image generator of OpenAI (figure 2). The conventional doctors' report could become interactive, understandable for patients and contains personalised images, for example, clinical courses or even physiotherapy exercises for the patient. These reports do not need to be static, but could be animated by AI, for example, by visually integrating digital biomarkers and individualised predictions of disease activity in arthritis.[11 12]

Without a doubt, the greatest risk in the use of chatbots in the field of rheumatology as elsewhere lies in the loneliness of patients who are particularly dependent on interpersonal exchange. I urge that we use the freed up time and resources wisely to get more interpersonal time to talk about therapy, nutrition, stress management, preventive care, ability to work, physical and psychological resources, vaccinations, etc.

CONCLUSION

Following speech and image recognition, text generation by large language models including chatbots is now entering medicine and thus also rheumatology. It will not completely change our clinical routine, but it will make information more easily available and save time, which we will hopefully use wisely. In particular, for chronic diseases such as RA, an immense amount of data accumulates in the course of a patient journey that we do not see and do not use. Chatbots will not be able to make

clinical decisions for now; for this, they are dependent on the relevant guidelines or studies that have been carried out. But they will be able, for example, help to design study protocols and carry out decentralised studies more easily. The basis of all this is the education around this new technology among both patients and healthcare professionals.

Acknowledgements I thank Dr. Chris Lovejoy for critically reviewing this manuscript.

Contributors TH is the only author of this article.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD

Thomas Hügle http://orcid.org/0000-0002-3276-9581

REFERENCES

1 Thorp HH. ChatGPT is fun, but not an author. *Science* 2023;379:313.
2 Graham F. Daily briefing: science urgently needs a plan for ChatGPT. *Nature* 2023.
3 Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* 2023;613:612.
4 Else H. Abstracts written by ChatGPT fool scientists. *Nature* 2023;613:423.
5 The Lancet Digital H. ChatGPT: friend or foe? *Lancet Digit Health* 2023;5.
6 Shen Y, Heacock L, Elias J, *et al*. ChatGPT and other large language models are double-edged swords. *Radiology* 2023;307:230163.
7 Chatterjee J, Dethlefs N. This new conversational AI model can be your friend, philosopher, and guide… and even your worst enemy. *Patterns* 2023;4:100676.
8 FDA. *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*.
9 Faes L, Wagner SK, Fu DJ, *et al*. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health* 2019;1:e232–42.
10 OpenAI API Community Forum. *Academic Reference Errors – ChatGPT*.
11 MedPage Today. *What Can ChatGPT Do For Your Practice?*
12 Hügle M, Walker U, Finckh A, *et al*. Personalized prediction of disease activity in patients with rheumatoid arthritis using an adaptive deep neural network. *PLOS One* 2021.