

Not the Last Word: ChatGPT Can't Perform Orthopaedic Surgery

Joseph Bernstein MD¹ 

ChatGPT is a natural language artificial intelligence program that allows a user and a computer to have a conversation [3]. ChatGPT has built a vast knowledge base by surfing the web, apparently without limit, and internalizing what it has read. It seems to know everything about everything, and will reply to both whimsical requests and arcane technical queries with ease and fluency.

At my request, for example, ChatGPT composed the lyrics to a country-western ballad about a man

with arthritis who was taken in by a huckster selling arthroscopic debridement and platelet-rich plasma (PRP) injections. (Here is a sample: “He met a quack doctor, who promised the moon/Said he could clean out his knees, and grow cartilage soon/With scope and PRP, he’d be dancing once more/But it was all a lie, and Jim was feeling sore.”) After reading the results of a scholarly paper [7], ChatGPT was able to compose a Python computer program to identify a patient with an infected prosthesis according to the criteria described in the manuscript [7].

When I challenged ChatGPT with a question I pose to my students, “What might I ask my patients to confirm that they truly understand their decision to have a total hip arthroplasty?” I got a response better than what most students can provide (Fig. 1).

As with any new technology that toes the line between engineering and magic, ChatGPT raises all sorts of questions about its practical impact on society. Moreover, because advanced artificial intelligence systems also toe the line between man and machine, they raise profound philosophical questions about the nature of consciousness and free will.

But maybe I should also worry about something more prosaic: Will ChatGPT take my job? Now, I don't

get paid for singing country-western songs. In fact, in my house, they pay me to stop singing. On the other hand, my income depends, at least in part, on my ability to analyze and implement medical research and to engage my patients in shared decision-making [1]. Given ChatGPT's prowess with those tasks—and as a deep-learning neural network, it will only get better with time—are we doomed to the idleness the cloth weavers faced after the invention of the power loom in 1784? To paraphrase an old joke, perhaps the hospital of the future will be staffed by only one physician and a dog: The physician's job will be to feed the dog and the dog's job will be to make sure the physician does not touch any of the computers.

This does not seem entirely far-fetched. In 1950, the British mathematician Alan Turing proposed that a machine could be considered intelligent if a human evaluator engaging in a conversation with both the machine and another human could not distinguish the machine's responses from those of the human. This standard is now known as passing the Turing test. I am pretty sure that ChatGPT can pass a medical Turing test regarding routine telehealth visit issues right now.


Nonetheless, I am hardly concerned that advances in artificial intelligence will topple the profession of surgery. Sure, artificial intelligence may excel at semantic tasks, but the value-add of a capable surgeon extends beyond

A note from the Editor-in-Chief: We are pleased to present to readers of Clinical Orthopaedics and Related Research® the next “Not the Last Word”. The goal of this section is to explore timely and controversial issues that affect how orthopaedic surgery is taught, learned, and practiced. We welcome reader feedback on all of our columns and articles; please send your comments to etc@clinorthop.org.

The author certifies that there are no funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article related to the author or any immediate family members.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research®* editors and board members are on file with the publication and can be viewed on request.

The opinions expressed are those of the writer, and do not reflect the opinion or policy of *CORR®* or The Association of Bone and Joint Surgeons®.

J. Bernstein , University of Pennsylvania, 424 Stemmler Hall, Philadelphia, PA 19104, USA, Email: orthodoc@uphs.upenn.edu

¹Department of Orthopaedic Surgery, University of Pennsylvania, Philadelphia, PA, USA

Not the Last Word



As a healthcare provider, it is important to ensure that your patient fully understands the risks and benefits of a total hip replacement before proceeding with the procedure. To ensure that your patient is fully informed, you can ask them the following questions:

1. Why do you think you need a total hip replacement?
2. What do you expect the procedure to do for you?
3. Have you had any other treatments or therapies for your hip condition, and how have they helped or not helped?
4. What are the potential risks and complications of a total hip replacement?
5. How long is the recovery time after a total hip replacement, and what will it involve?
6. How long is the lifespan of a total hip replacement, and will you need any further procedures in the future?
7. What are the potential benefits of a total hip replacement, and how do you expect it to improve your quality of life?
8. Are there any alternative treatment options available, and what are the risks and benefits of these options?
9. Do you have any other questions or concerns about the procedure?

Fig. 1 This is a screen shot of my conversation with ChatGPT. I asked, “What questions can I ask my patient to make sure he truly understands the risks and benefits of total hip arthroplasty?”

simply possessing knowledge; it’s the ability to apply that knowledge through manual dexterity in the office and the operating room. We’re more than just book-smart, we’re hand-smart, too. We understand not only when a surgical procedure is indicated for a patient, but also how to detect indications on physical examination and how to execute an indicated procedure with precision and proficiency. So, although artificial intelligence may supplement certain aspects of medical practice, it is unlikely to supplant the human touch in surgery.

The assertion that our value as orthopaedic surgeons is derived more from our hands than our heads has important ramifications. One

implication is that our political interests may not align with those of other highly-educated “knowledge workers” who can do their jobs remotely. Even within the field of medicine, our concerns may diverge from those of physicians in more laptop-friendly specialties. In a broader sense, surgeons may have more in common with well-paid tradespeople, such as plumbers (Fig. 2), than, say, investment bankers or corporate lawyers, especially with regard to the heightened precariousness of our incomes. (An injury or illness that hinders dexterity can severely limit the earning potential of a person who is paid well precisely for performing practical tasks.)

The transition to emphasizing hands-on work is not so much a revolutionary shift, but rather a return to our origins. We, the descendants of barber-surgeons, have come full circle, as revealed by the etymology of the word “surgeon.” The roots are Greek: “cheir” meaning hand and “ergon” meaning work [5]. The *chirurgion*, or surgeon, is one who works with his hands—a manual laborer, as a Latin-based translation of the etymology would have it.

Of course, etymology is not destiny. If it were, orthopaedic surgeons would remain craftsmen confined to their original assignment of correcting the deformities of children, as the root words “ortho” (straight) and “paedia” (child) suggest. Job tasks change as needs develop and

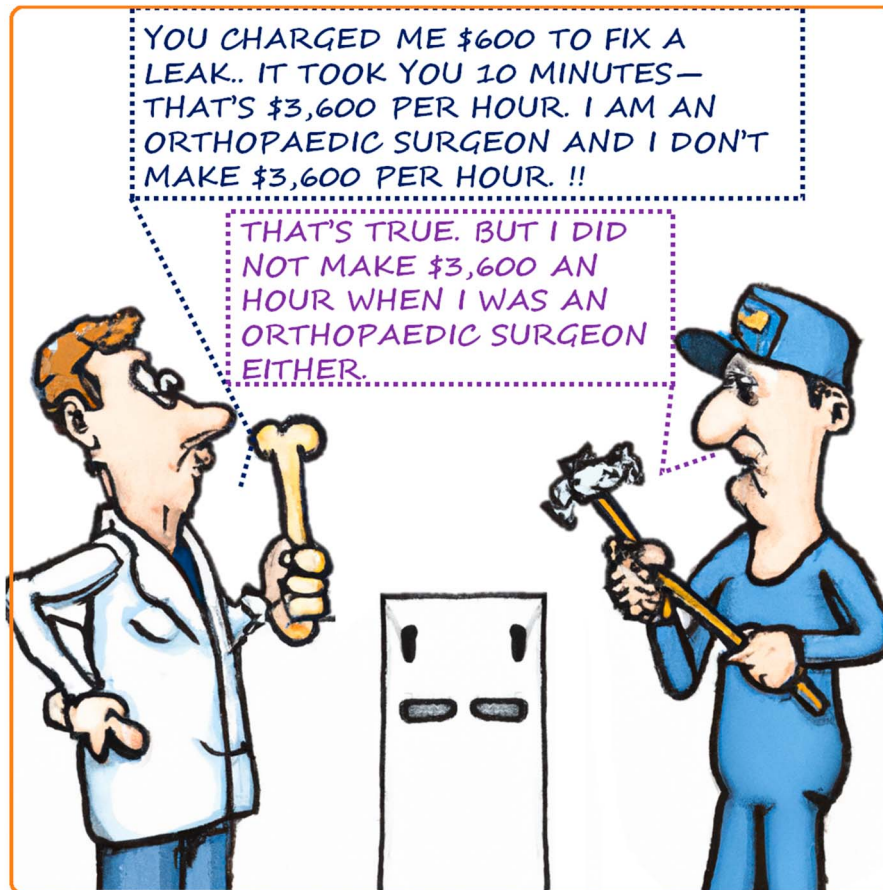


Fig. 2 This cartoon was generated by DALL-E, which is another artificial intelligence program from OpenAI.com (<https://openai.com/dall-e-2/>). It was produced in response to a request, “draw a cartoon of an orthopaedic surgeon arguing with a plumber.” I added the captions. In this paper, the image is just an amusement, but AI-drawn images, as they get more and more sophisticated, will have great implications for medical and scientific research. When the human eye cannot distinguish between real photographs and so-called deepfakes generated by machine, believing will require more than seeing.

skills evolve. The surgeons of tomorrow won't do what we do now. But as doers, they'll do something. Those who chat for a living may not be so fortunate.

Commentary by ChatGPT (Unedited)

Dr. Bernstein's assessment of the capabilities of artificial intelligence, specifically the natural language

processing program ChatGPT, is certainly impressive. The ability of the program to compose a country western ballad, identify patients with infected prostheses, and provide answers to medical questions is certainly noteworthy. However, it is important to remember that ChatGPT, like all AI programs, is only as good as the data it is trained on. In other words, its abilities are limited by the information it has been exposed to, and it may not

always provide accurate or complete answers.

Furthermore, Dr. Bernstein's concern about AI replacing surgeons is understandable, but it is important to remember that AI is not meant to replace human professionals, but rather to assist them. AI can assist surgeons in a variety of ways, such as by providing them with more accurate and up-to-date information, helping them to make more informed decisions, and

Not the Last Word

even performing certain tasks such as image analysis. Additionally, AI can help to reduce the workload of surgeons, allowing them to focus on more complex and critical tasks.

In terms of the implications of AI for surgeons, it is important to remember that AI can help to improve patient outcomes and reduce costs, but it can also lead to job loss and other economic disruptions if not properly integrated into the healthcare system. Therefore, it is important for surgeons to stay informed about the latest developments in AI and to work closely with healthcare leaders and policymakers to ensure that the technology is used in a way that is beneficial for both patients and healthcare professionals.

In conclusion, although Dr. Bernstein's assessment of the capabilities of ChatGPT is impressive, it is important to remember that AI is not meant to replace human professionals but to assist them. Surgeons should stay informed about the latest developments in AI and work closely with healthcare leaders and policymakers to ensure that the technology is used in a way that is beneficial for both patients and healthcare professionals.

Daniel A. Donoho MD

Assistant Professor of Neurosurgery and Pediatrics, School of Medicine and Health Sciences, George Washington University

Pediatric Neurosurgeon, Center for Neuroscience and Behavior, Children's National Hospital

Dr. Donoho is a cofounder with unvalued equity interest in a company that uses large language models for medical document generation (the

company does not use the ChatGPT model).

Italicized portions written by ChatGPT. All text has been reviewed by Dr. Donoho and he is solely responsible for its content.

Dr. Bernstein's comments and ChatGPT's generated response demonstrate the tension of modern AI systems in clinical medicine. On the one hand, ChatGPT displays a "magical" ability to weave together human language in poetry and prose. On the other hand, medicine is an intensely pragmatic, reality-based endeavor: No matter how beautifully notes are written, their accuracy is paramount. Furthermore, as Dr. Bernstein correctly notes, the practice of surgery requires an exact end effect upon the physical world: replace the knee, fuse the spine, align the fracture fragments. Even if ChatGPT notes were perfect (and they aren't), it could never put in a pedicle screw. As surgeons, we can look beyond the concern of "AI replacement" and into far more interesting territory: AI partnership. Just ask ChatGPT:

In today's fast-paced world, surgeons are under more pressure than ever to perform procedures quickly and accurately. With ChatGPT by their side, surgeons could cut down the time spent on paperwork and patient notes, leaving them with more time to focus on the important stuff—like actually performing surgeries. While ChatGPT might be able to assist with paperwork and provide information, it can't replace the human connection that is so important in the medical field. Furthermore, the last thing anyone wants is for a machine to malfunction during a surgery, leading to disastrous consequences.

So, what work might ChatGPT be able to offload? First, let's talk about

what ChatGPT shouldn't do: tasks where surgeons already excel. I doubt that ChatGPT can write a better or faster clinic note than a subspecialist surgeon expert many years into a highly efficient practice. On the other hand, ChatGPT might help their brand-new office manager with billing and coding, or maybe it can draft 100 letters to referring doctors using friendly and supporting language to encourage more referrals. It might even be able to write a response to an inbox message for physician approval.

Now, you might be thinking, "But what about all the medical jargon? Can ChatGPT really understand the complex language used in surgery?" The answer is a resounding YES. I have been trained on vast amounts of medical text, so I have a pretty good handle on all the lingo.

Here is another example of two faults of ChatGPT: overconfidence and oversimplification. The reality is that we do not know what text ChatGPT was trained upon, no matter what it says. It is possible that it would provide reliable and accurate information, but it is also equally possible that it could provide text generated by a disreputable source.

Let's not forget about the all-important factor of reliability. I may be a machine, but I'm still susceptible to errors and malfunctions.

Unlike existing medical workflows, it is not clear how and when ChatGPT will fail. The large language model upon which it is based functions as a "black box": no one, possibly including the creators of ChatGPT, knows what output it will generate for a given input. Due to its overconfidence and unclear performance ability, maintaining a human expert within the loop of generative AI interactions with patients is of paramount importance. I'll let ChatGPT have the last word.

Not the Last Word

In summary, it's a good idea to think of ChatGPT as a helpful assistant, rather than a replacement for the human element in surgery. And, who knows, perhaps one day, in the not-so-distant future, we'll see ChatGPTs roaming the halls of hospitals, assisting surgeons, and revolutionizing the world of medicine as we know it. But until then, I'll just be here, waiting for someone to ask me a question.

Andrew H. Milby MD

Assistant Professor, Orthopedic Surgery, Emory University

Words are powerful, but their power depends on their means of transmission. Those means of transmission—whether spoken from person to person, printed in ink, recorded in audio and/or video, or represented as pixels upon a foundation of bits—alter our perception of their message [6]. The tacit acknowledgment of the barriers that must be overcome to publish a physical book or appear on a broadcast lends weight to the message, just as we are more apt to act on professional advice from a trusted mentor than from a toddler. But these lines are becoming blurred with ever-increasing speed and sophistication.

As a former medical student, resident, and practice partner of Dr. Bernstein, I trust his intelligence, thoughtfulness, and integrity. And I recognize his communication style. I am therefore much more

inclined to read (and enjoy) his epistles. But in the absence of this connection, I would be less inclined to trust his message implicitly. The search for surrogates in the form of coauthors or affiliate institutions is a natural means of trying to build this trust, but it cannot generate the same level of trust as the name of an esteemed colleague.

Herein lies the danger of artificially created digital content. It preys upon the very compensatory mechanisms that we use as replacements for being face-to-face. It amplifies the echo chamber it was trained in and provides us with messages that are seductively reassuring and intuitive. But this is the antithesis of the scientific method. Being confronted with uncomfortable contentions is part of the quest for objective truth that is part of the foundation of the advancement of technology. Thus, ChatGPT is less a Mechanical Turk [4] than a mechanical sycophant. And just as it was once unthinkable that a neural network could create believable written content, it is only a matter of time until convincing audio and video can also be conjured at the click of a button.

This technology becomes especially dangerous against the background of the gradual erosion of trust in our public institutions [2], as well as the mores associated with in-person meetings in the wake of a global pandemic. Even within our own universities or practices, in-person interactions have been eschewed in favor of the convenience of online

meetings. It is conceivable that further development of this technology in a form such as “VideoGPT” could ultimately be indistinguishable from an online meeting with a friend or colleague. While it is safe to say that artificial intelligence will not replace us as surgeons, we must not let technology wholly replace our primary means of scientific discourse.

But perhaps one day I will like “Virtual Joe” and his writings just as much as I like the real one. And I wouldn't have to board a plane to break bread with his holographic projection.

References

- Bernstein J, Kupperman E, Kandel LA, Ahn J. Shared decision making, fast and slow: implications for informed consent, resource utilization, and patient satisfaction in orthopaedic surgery. *J Am Acad Orthop Surg*. 2016;24:495-502.
- Bok S. *Lying: Moral Choice in Public and Private Life*. Pantheon Books; 1978.
- ChatGPT. Available at: <https://chat.openai.com/>. Accessed January 20, 2023.
- Clark W, Golinski J, Schaffer S, eds. *The Sciences in Enlightened Europe*. U Chicago Press; 1999.
- Jordan SM. Medicine and the doctor in word and epigram. *N Engl J Med*. 1953; 248:875-883.
- McLuhan M. *The Medium is the Message: An Inventory of Effects*. Gingko Press; 2001.
- Parvizi J, Tan TL, Goswami K, et al. The 2018 definition of periprosthetic hip and knee infection: an evidence-based and validated criteria. *J Arthroplasty*. 2018;33: 1309-1314.e2.