Value-based Healthcare

Jayakumar, Prakash; Oude Nijhuis, Koen D.; Oosterhoff, Jacobien H.F.; Bozic, Kevin J.

Published in:
Clinical Orthopaedics and Related Research

DOI:
10.1097/CORR.0000000000002854

Publication date:
2023

Clin Orthop Relat Res (2023) 481:1890-1894
DOI 10.1097/CORR.0000000000002854

Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

Value-based Healthcare

Published online: 7 September 2023

Value-based Healthcare: Can Generative Artificial Intelligence and Large Language Models be a Catalyst for Value-based Healthcare?

Prakash Jayakumar MD, PhD[1], Koen D. Oude Nijhuis MD[1,2], Jacobien H. F. Oosterhoff MD, PhD[3], Kevin J. Bozic MD, MBA[1]

There has been an explosion of interest in generative artificial intelligence (generative AI)—a class of machine-learning models that are trained using large volumes of text, audio, image, or video data to generate new, natural, human-like content (Fig. 1) [2]. Large language models (LLMs), commonly known as chatbots (such as the OpenAI company's generative pretrained transformer "ChatGPT"), are a form of generative AI, trained using a vast number of text data parameters to create algorithms that perform numerous language-specific tasks. ChatGPT v3.5, released on November 30, 2022, is the fastest-growing consumer application in history [5], and its current iteration, ChatGPT-4, leads a growing pack of LLMs, able to process sophisticated instructions to generate increasingly creative and accurate outputs from multimodal data.

Widespread access to LLMs and their rapid expansion in healthcare has signaled the need for urgent attention. Amid the hope and hype, we'll take a pragmatic view and envision some high-value opportunities for generative AI, outline the potential threats associated with them—some of which were identified by the Center of AI Safety [3]—and suggest strategies to overcome those threats.

Value Opportunity #1: Improving Patient Experience

A common application of LLMs is the ability to freely submit queries, such as symptoms causing concern, to generate possible diagnoses and treatment suggestions [1]. Rapid access to such health information, especially for patients with wide-ranging levels of health literacy, can be beneficial, particularly in health systems where access to primary care is limited or underfunded [11]. LLMs may further act as virtual assistants, linking health advice to automated appointment scheduling, thereby enhancing patient experience and continuity of care.

Potential Threat #1: Medical Misinformation

Google's LLM "Bard" falsely stated in an early promotional advert by the company that the James Webb Space Telescope was used to take the first pictures of a planet outside our solar system when these were in fact taken by the European Southern Observatory's Telescope in 2004 [4]. Further, there are several instances where seemingly credible scientific citations have been

A note from the Editor-in-Chief: We are pleased to present to readers of Clinical Orthopaedics and Related Research® *the latest* "Value-based Healthcare" *column (formerly* "Orthopaedic Healthcare Worldwide"*).* "Value-based Healthcare" *explores strategies to enhance the value of musculoskeletal care by improving health outcomes and reducing the overall cost of care delivery. We welcome reader feedback on all of our columns and articles; please send your comments to eic@clinorthop.org.*

One of the authors (KJB) certifies receipt of personal payments or benefits, during the study period, in an amount of less than USD 10,000 from The Centers for Medicare & Medicaid Services.
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 writers, and do not reflect the opinion or policy of *CORR®* or The Association of Bone and Joint Surgeons®.

K. J. Bozic ✉, Dell Medical School at the University of Texas at Austin, 1601 Trinity Street, Building B, Austin, TX 78712, USA, Email: kevin.bozic@austin.utexas.edu

[1]Department of Surgery and Perioperative Care, Dell Medical School at the University of Texas at Austin, Austin, TX, USA

[2]Department of Orthopedic Surgery, The University of Groningen, Groningen, the Netherlands

[3]Department of Engineering Systems and Services, Faculty Technology Policy and Management, Delft University of Technology, Delft, the Netherlands

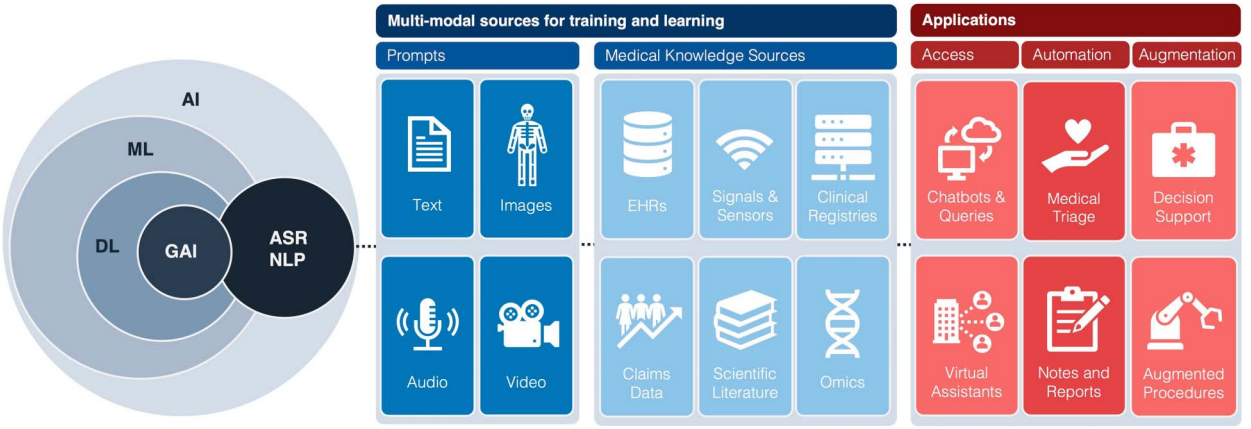

Fig. 1 A schematic showing the position of generative artificial intelligence (AI) and large language models among AI technologies alongside different forms of data prompts to generate new content. ML = machine learning; DL = deep learning; GAI = generative AI; ASR = automatic speech recognition; NLP = natural language processing (including large language models). A color image accompanies the online version of this article.

fabricated in response to clinical queries [8]. Such plausible-sounding but incorrect, out-of-date, or simply conjured up material, known as "AI hallucinations," can fuel misinformation, inaccurate information transfer, and less effective decision-making. This phenomenon is challenging to resolve given the source of truth may often be lost during model training.

While it is clear that generative AI has instant creative power and learning potential, the risk of intensified distress and unhelpful thinking through medical misinformation underscores some vulnerabilities behind this technology and a demand for human stewardship along the patient's healthcare experience in order to ensure safety, trust, and quality [6].

Value Opportunity #2: Improving Patient Outcomes

Generative AI and LLMs also have the potential to improve patient outcomes by learning from large volumes of structured data (such as payer claims-based data and clinical registries) and unstructured data (like clinic notes in electronic medical records [EMRs]) to provide personalized predictions that can aid clinical decision support and shared decision-making. For instance, population-level data from the Global Burden of Disease Study could be used to provide personalized disease assessments and multimodal, coordinated management of common conditions alongside preventative health strategies [14]. Companies such as Bionic Health are using LLMs in tandem with physicians to provide such evaluations alongside diagnostic services. Others are also integrating this technology with data received from devices in lived environments (like the Apple watch and Amazon Alexa) and devices in hospital settings for monitoring and triggering clinical alerts. Further, LLMs have been developed to address mental health concerns by recognizing symptoms of depression from speech alone, which is something clinicians can easily miss [15]. Mindsets explain a substantial amount of the variation in symptom intensity and level of capability among musculoskeletal patients. Thus, orthopaedic surgeons might consider greater access to mental health support and risk stratification of psychological factors a welcome addition to their practices, even if it's being done by a computer.

Potential Threat #2: The Black Box and Lack of Transparency

Careful selection, weighting, and handling of data parameters in generative AI and LLMs is critical if we are to apply algorithms for decision support at scale. However, complex neural networks, especially those involving models built on expansive datasets, become black boxes lacking transparency and generalizability. Replicability, standardization, and correction of biases and errors within such models becomes increasingly challenging as datasets grow exponentially. This unbridled growth can compromise trust in the technology to deliver accurate outputs that are able to improve patient outcomes. Furthermore, generative AI and LLMs involve some element of supervised learning, raising questions about who is performing and supervising that learning.

Value Opportunity #3: Improving Health System Costs

Generative AI is accelerating the performance and usability of computers as they become exponentially faster and more powerful. This may result in substantial improvements in efficiency and decision-making, which can also translate to reduced costs. For instance, the automation of routine administrative tasks, clinical tasks, and medical triage by translating voice and text prompts into operational actions can substantially decrease the time it takes to perform them, lower the associated cost, increase productivity, and potentially raise work enjoyment [10].

Potential Threat #3: Cost Crisis

Increasing utilization and preprocessing of massive datasets to develop generative AI and LLMs can be expensive, both in terms computational processing power and power consumption [9], and those costs will only increase as demand for the technology rises. Further, while generative AI and LLMs promise to improve performance and efficiency, the technology may paradoxically perpetuate existing problems in the US healthcare system. For instance, if models are taught from data within the current system, they may drive overutilization of expensive, low-value tests and treatments, like MRIs and injectable biologics for knee osteoarthritis, which remain commonplace in orthopaedic practices.

Value Opportunity #4: Improving Care Team Well-being

Clinicians spend a substantial amount of time performing repetitive, non-patient-facing activities that involve navigating multiple systems (EMRs, imaging software, billing platforms) [13]. Generative AI can reduce this burden through automation, thereby reducing fatigue, cognitive burden, and freeing-up doctors' time for their patients, care teams, and personal lives. The Microsoft Nuance system aims to automate clinical documentation and complete administrative tasks within the EMR by "listening" to clinician-patient conversations using voice recognition software during medical visits. Using advanced reasoning and natural language capabilities, the platform transcribes and summarizes notes in seconds using voice-enabled digital scribes integrated with ChatGPT-4.

Clinician experience may be further enhanced through LLM-driven clinical decision support functions within EMRs, such as differential diagnosis generation, risk prediction, treatment planning, and medical error prompts. These tools may also be useful for monitoring patients via other data sources like video/audio-based sensors. Administrative and clinical benefits for care teams across different settings may increase exponentially through generative AI applied to data from networks in different locations that are connected to each other through shared technological resources managed by a central framework called a federated system.

Potential Threat #4: Enfeeblement

As generative AI solutions take on a growing number of clinical tasks with increasing proficiency, the dependence on technology, which some have termed "enfeeblement," will undoubtedly increase. This may potentially reduce the number of opportunities for clinicians to learn and hone their skills, instead increasing their reliance on digital systems. This might also result in concerns about accountability when medical errors occur.

Value Opportunity #5: Improving Health Equity

Responsible generative AI and LLM model design that addresses the various aspects of ethical AI—including privacy, bias, transparency, trustworthiness, and accountability—has the potential to alleviate health disparities and provide more equitable care. Those from minority populations, underserved communities, and individuals with mobility issues may find greater access to tailored health information and guidance on health issues via generative AI. Further, AI-generated synthetic data can also level the information playing field in equitable clinical research by making access to secondary use of data easier.

Potential Threat #5: Amplification of Biases

Amplification of intrinsic biases and errors can occur within large datasets, such as automation bias (a tendency to favor results generated by automated systems over those generated by non-automated systems, irrespective of error rates) and selection bias (where data selected is not reflective of the population). In healthcare, such biases can lead to potentially harmful discrimination against some patient populations [6]. For instance, current generative AI models are built on data that are largely based on Western populations [15]. Further, only those with resources to access the internet can experience these models, thereby perpetuating biases as their inputs will not be included in future model development. Experts also predict consolidation of generative AI-related

activity and its development being led by fewer and fewer hands alongside greater surveillance and censorship. This trend could exacerbate rather than alleviate disparities. For instance, generative AI platforms automatically seeking approvals for health plans on behalf of doctors may overlook rather than address underlying patient needs or access to care issues for vulnerable populations.

Guiding Principles for a Generative AI Strategy

Principle 1. Utilization of High Quality, High Volume, Secure Data

Accurate, appropriate, and contextually relevant outputs of generative AI demand access to high-quality, high-volume, diverse datasets that are defined by type, source, and timeframes over which data were collected. Strong data privacy and security safeguards are required around data provided by patients following informed consent and should be a requisite of future federated datasets and generative AI solutions. Patients place substantial trust in health systems to carry and utilize their personal health information and therefore should have their data adequately protected.

Principle 2. Development and Regulation of Generative AI

Governments, companies, health systems, medical professional societies, and academia need to act fast and craft a comprehensive strategy for AI regulation, especially considering the staggering pace of advancement in this technology.

In the US, federal regulation is in its early stages. Initiatives such as the National Institute of Standards and Technology (NIST) AI Risk Management Framework, blueprint for an AI Bill of Rights, and governance measures by the Federal Trade Commission (FTC) are a start. Regulations should guide documentation and disclosure of AI methods and data sources, safety pre- and post-testing, model licensing and auditing, and include risk-based regulation around use-cases, such as that defined in the European Union (EU) AI Act, rather than blanket regulation encompassing the whole technology. Global cooperation around AI governance should also be driven through forums such as the Organization for Economic Cooperation and Development and Global Partnership for AI and emulating strong legislation such as the EU General Data Protection and Regulation laws.

Generative AI companies have a responsibility to abide by a foundational set of findable, accessible, interoperable, and reusable (FAIR) principles, originally developed in 2016 as a requisite for proper data management and stewardship and since adapted for AI models [7]. Regulation should mandate internal governance processes, mitigate harm by limiting or removing personal data from training data sets, and promote stringent internal testing prior to release, alongside independent validation and human feedback, so models can be improved. Companies like Google are investing heavily in testing their generative AI models, recognizing significant trust should be built with industry as this technology becomes deep-rooted in society.

Principle 3. Implementation of Legal and Ethical Guidelines

Most users are unaware that their queries will be used to train further models and that their personal information can be relatively easily linked to medical information and accessed by companies without the necessary safeguards. Regulatory frameworks set by government agencies, such as the NIST and the FTC, and professional societies around generative AI should include ethical and legal guardrails around access to data and alignment of models with the best interests of patients. These guardrails should be developed in collaboration with software developers, healthcare professionals, ethicists, legal experts, and patient advocates. Such frameworks should include penalties for noncompliance and be set up to respond to complaints about model safety and accuracy in relation to the level and context of use. For instance, if LLMs are used to augment rather than replace professional involvement, this falls under the heavily regulated legal framework and well-defined aspects of the clinician-patient relationship that align with values of competency, trust, and patient autonomy; as such, LLMs in this context might be regulated accordingly. Advice, whether good or bad, generated by LLMs is no different than information from other sources. Digital watermarks or disclosures specifying the involvement of AI, similar to those used in other formats of health information transfer, provide one solution. Either way, the clinician is ultimately responsible, and therefore liable, for the outcome. Thus, it is imperative that clinicians understand how generative AI works, become familiar with validated tools, and recognize the limitations of this technology.

Far fewer guardrails exist around patient-facing generative AI and LLM applications in which a professional is removed from the

equation. Instead, patients may trust the human-like outputs of the generative AI model, making patients more vulnerable. This was observed when a company providing emotional support chat services for people in distress transitioned its human counselors to AI-generated messages via a chatbot without the knowledge of their customers [6]. The lack of regulation and malpractice liability coverage should halt further autonomous applications of generative AI and LLMs in healthcare until an adequate level of professional oversight can be implemented. Currently, most if not all consumer-facing generative AI applications are not HIPAA compliant. The FDA should hold generative AI developers accountable for false medical claims and work with the FTC in cases of medical misinformation as deceptive business practices. Regardless, both patients and clinicians should be educated about the dangers of using unregulated generative AI applications outside the clinician-patient relationship.

Generative AI and LLMs are expanding the reach of AI technology within healthcare and our daily lives. While such technology is poised to enable a variety of value-generating opportunities, stakeholders should be aware of potential threats and supportive strategies for data quality, model development and regulation, and legal and ethical guardrails.

Acknowledgments We thank Professor Job Doornberg, Professor of Orthopaedics, Chairman of the AI Think Tank at the University Medical Center Groningen, and expert Clinical Application of AI at the Data Science Center in Health (DASH), and Dr. Joseph Schwab, Associate Professor of Orthopaedic Surgery, Harvard Medical School, for their comprehensive review and guidance on this column.

References

1. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med.* 2011;155:97-107.
2. Bloomberg Intelligence. Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds. *Bloomberg.* Published June 1, 2023. Available at: https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/. Accessed July 19, 2023.
3. Center for AI Safety. Available at: https://www.safe.ai. Accessed August 28, 2023.
4. Coulter M, Bensinger G. Alphabet shares dive after Google AI chatbot Bard flubs answer in ad. *Reuters.* Published February 8, 2023. Available at: https://www.reuters.com/technology/google-ai-chatbot-bard-offers-inaccurate-information-company-ad-2023-02-08/. Accessed July 27, 2023.
5. Gordon C. ChatGPT is the fastest growing app in the history of web applications. *Forbes*. Published February 2, 2023. Available at: https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applications/. Accessed July 19, 2023.
6. Haupt CE, Marks M. AI-generated medical advice—GPT and beyond. *JAMA*. 2023;329:1349-1350.
7. Huerta EA, Blaiszik B, Brinson LC, et al. FAIR for AI: an interdisciplinary and international community building perspective. *Sci Data*. 2023;10:487.
8. Leopold SS, Haddad FS, Sandell LJ, Swiontkowski M. Editorial: artificial intelligence applications and scholarly publication in orthopaedic surgery. *Clin Orthop Relat Res.* 2023;481:1055-1056.
9. McGregor J. Generative AI breaks the data center: data center infrastructure and operating costs projected to increase to over $76 billion by 2028. *Forbes*. Published May 12, 2023. Available at: https://www.forbes.com/sites/tiriasresearch/2023/05/12/generative-ai-breaks-the-data-center-data-center-infrastructure-and-operating-costs-projected-to-increase-to-over-76-billion-by-2028. Accessed August 8, 2023.
10. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5:e107-108.
11. Schoen C, Osborn R, Squires D, Doty MM. Access, affordability, and insurance complexity are often worse in the United States compared to ten other countries. *Health Aff (Millwood).* 2013;32:2205-2215.
12. Sham AH, Aktas K, Rizhinashvili D, et al. Ethical AI in facial expression analysis: racial bias. *SIViP*. 2023;17:399-406.
13. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med.* 2016;165:753-760.
14. Temsah MH, Jamal A, Aljamaan F, Al-Tawfiq JA, Al-Eyadhy A. ChatGPT-4 and the global burden of disease study: advancing personalized healthcare through artificial intelligence in clinical and translational medicine. *Cureus*. 2023;15:e39384.
15. Tian H, Zhu Z, Jing X. Deep learning for depression recognition from speech. *Mobile Networks and Applications.* 2023:1-6.