

## Review

# A Primer on Large Language Models (LLMs) and ChatGPT for Cardiovascular Healthcare Professionals

Muneeb Ahmed, MD, BHSc,<sup>a</sup> Jeffrey Lam, MD,<sup>b</sup> Alexander Chow, IBDP,<sup>c</sup> and Chi-Ming Chow, MDCM, MSc, FRCPC, FCCS<sup>a</sup>

<sup>a</sup> Department of Medicine, University of Toronto, Toronto, Ontario, Canada

<sup>b</sup> Division of Cardiology, Department of Medicine, Queen's University, Kingston, Ontario, Canada

<sup>c</sup> Brown University, Providence, Rhode Island, USA

## ABSTRACT

Generative artificial intelligence (AI), particularly large language models (LLMs), such as ChatGPT, is transforming healthcare by offering novel ways to synthesize and communicate medical knowledge. This development is especially relevant in cardiology, as patient education, clinical decision-making, and administrative workflows play pivotal roles in this area. ChatGPT, originally built on GPT-3 and refined into GPT-4, can simplify complex cardiology literature, translate technical explanations into plain language, and address questions across different linguistic backgrounds. Studies show that although ChatGPT demonstrates considerable promise in performing text-based tasks—ranging from passing portions of the European Exam in Core Cardiology to creating patient-friendly educational materials—its inability to interpret images remains a major limitation. Meanwhile, concerns around false information, data bias, and ethical issues highlight the need for careful oversight. Future directions include integrating LLMs with computer-vision modules for image-based diagnostics and combining unstructured patient data to improve risk prediction and phenotyping. Social-media research suggests that chatbots sometimes provide more-empathetic responses than do physicians, underscoring both their potential advantages and complexities. LLM-based tools can also generate letters for insurance prior authorizations or appeals, helping reduce administrative burden. New multimodal approaches, such as ChatGPT Vision, have the potential to enable direct image processing, although clinical validation of this function is yet to be established. The judicious integration of ChatGPT and other LLMs into cardiology requires ongoing validation, robust regulatory frameworks, and strong ethical guidelines to ensure patient privacy, avoid misinformation, and promote equitable healthcare de-

## RÉSUMÉ

L'intelligence artificielle (IA) générative, en particulier les grands modèles de langage (LLM) tels que ChatGPT, transforme les soins de santé en offrant de nouveaux moyens de synthétiser et de communiquer les connaissances médicales. Ceci est particulièrement pertinent en cardiologie, où l'éducation des patients, la prise de décision clinique et les flux de travail administratifs jouent un rôle central. ChatGPT, basé à l'origine sur GPT-3 et affiné dans GPT-4, peut simplifier la littérature complexe en cardiologie, traduire des explications techniques en langage clair et répondre à des questions provenant de différents contextes linguistiques. Les études montrent que si ChatGPT démontre un potentiel très prometteur pour les tâches basées sur le texte - qui vont de la réussite de certaines parties de l'examen européen de cardiologie de base à la création de matériel éducatif accessible aux patients - son incapacité à interpréter les images reste une limitation majeure. Par ailleurs, les inquiétudes concernant les renseignements faux, les biais de données et les questions éthiques soulignent la nécessité d'une supervision attentive. Les orientations futures comprennent l'intégration des LLM avec des modules de vision par ordinateur pour les diagnostics basés sur l'image et la combinaison de données non structurées sur les patients pour améliorer la prédiction des risques et le phénotypage. Les recherches sur les médias sociaux suggèrent que les agents conversationnels peuvent parfois fournir des réponses plus empathiques que les médecins, révélant des opportunités et les complications inhérentes. Pour les tâches administratives, les outils basés sur les LLM peuvent générer des lettres pour les demandes d'autorisation préalable ou des appels, contribuant ainsi à réduire la charge administrative. De nouvelles approches multimodales, telles que « ChatGPT Vision », pourraient à

Received for publication November 29, 2024. Accepted February 17, 2025.

Corresponding author: Chi-Ming Chow, University of Toronto, Echocardiography Lab, St. Michael's Hospital, 30 Bond St, Toronto, Ontario M5B 1W8, Canada. Tel.: +1 416 864-3024; fax: +1 416 864-5330.

E-mail: [Chi-Ming.Chow@unityhealth.to](mailto:Chi-Ming.Chow@unityhealth.to)

✉ [@muneebahmed1a](mailto:@muneebahmed1a), [@ChiMingChow1](mailto:@ChiMingChow1)

See page 665 for disclosure information.

Cardiovascular healthcare involves intricate diagnostic evaluations, extensive patient education, and substantial administrative demands. Recent advances in generative AI—particularly large language models (LLMs)—promise to address many of these challenges. ChatGPT (OpenAI, San Francisco, CA) exemplifies how LLMs can generate human-like responses by recognizing patterns across billions of text tokens from online

livery. This review aims to provide a primer on LLMs for cardiovascular professionals, summarizing key applications, current limitations, and prospects in this rapidly evolving field of digital health.

sources. Released in late 2022, ChatGPT quickly captured the public's attention for its ability to engage in a conversational question and answer session, correct mistakes, and recall prior interactions.<sup>1</sup>

Although ChatGPT and other LLMs offer transformative potential, they also present notable shortcomings. The process of directly interpreting medical images, such as electrocardiograms (ECGs) and echocardiograms, in their raw form,<sup>2</sup> has limitations, and the images can occasionally present “hallucinations”—that is, false information. Moreover, healthcare professionals must ensure patient privacy when inputting data into third-party AI systems.<sup>3</sup> This article provides a primer on ChatGPT and other LLMs, focusing specifically on their applications, limitations, and ethical implications in cardiovascular medicine. The goal of the article is to introduce cardiovascular professionals to the potential use of LLMs in their field.

## Generative AI and How LLMs Work

Generative AI uses machine-learning architectures that produce new content—text, audio, or images—by analyzing patterns from large datasets.<sup>4</sup> LLMs within this category specialize in generating text based on context learned during training.<sup>5</sup>

A breakthrough came with the introduction of transformer architecture, which processes text in parallel (rather than sequentially, as did older recurrent neural networks).<sup>6</sup> This innovation, referred to as “self-attention,” allows the model to weigh the relevance of different words or tokens in the same input, dramatically increasing both accuracy and training efficiency.

Figure 1 illustrates how LLMs tokenize an input sentence and apply attention layers to predict the most-plausible next token. ChatGPT initially was powered by GPT-3, which had 175 billion parameters.<sup>7</sup> Subsequent refinements and the incorporation of human feedback led to GPT-3.5, and then GPT-4, with enhanced reasoning capabilities.<sup>8</sup> Notably, GPT-4 remains primarily a text-based model that cannot interpret raw images.<sup>2</sup>

## Other LLMs in Healthcare

Although ChatGPT's popularity has propelled it to the forefront, multiple other LLMs are poised to influence healthcare, such as the following:

- Med-PaLM 2 (Google)—a variant of PaLM 2 fine-tuned for medical tasks, alleged to perform well on

l'avenir permettre un traitement direct des images, bien que la validation clinique reste à établir. En fin de compte, l'intégration judicieuse de ChatGPT et d'autres LLM en cardiologie nécessitera une validation continue, des cadres réglementaires solides et des lignes directrices éthiques strictes pour garantir la confidentialité des patients, éviter la désinformation et promouvoir une prestation de soins de santé équitable. Cette étude a pour but de fournir aux professionnels de la santé cardiovasculaire une introduction aux LLM, en résumant les applications clés, les limites actuelles et les perspectives dans ce domaine de la santé numérique qui évolue rapidement.

the tasks of clinical reasoning and question answering<sup>9</sup>;

- Llama 2 (Meta)—an open-source LLM available for research and commercial use, offering customization for specific domains (eg, cardiology)<sup>10</sup>; and
- Claude (Anthropic)—a model that focuses on “safer” outputs by limiting harmful or hallucinated text, potentially reducing misinformation.

Comparative studies are ongoing, but these alternative models share a similar transformer backbone and can be fine-tuned for use in specific medical specialties.

## Applications in Cardiovascular Medicine (Fig. 2)

### Simplifying medical literature and patient education

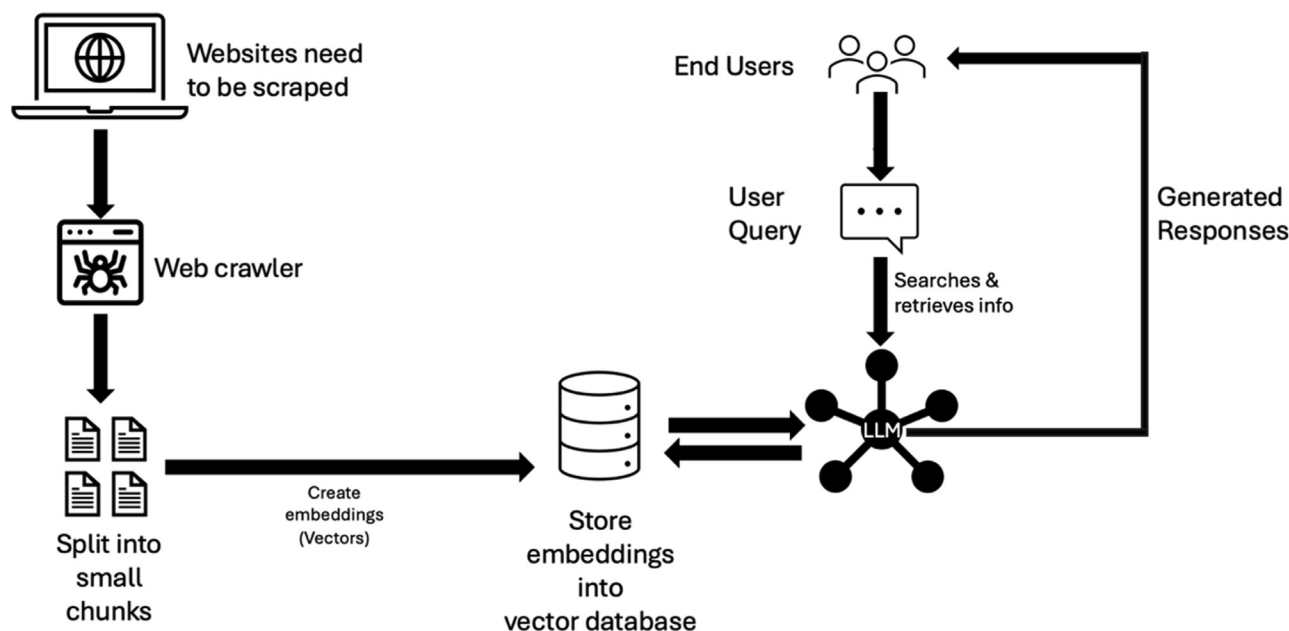
Medical literature frequently exceeds the average reading and comprehension level of adults,<sup>11</sup> including those with lower socioeconomic status. ChatGPT can convert complex cardiology texts—such as guidelines on heart failure or cardiac catheterization—into plain language without omitting essential content.<sup>12,13</sup> Such simplified explanations can enhance patient health literacy, potentially improving adherence and outcomes.<sup>14</sup>

ChatGPT's ability to provide patient-centred education has been studied for chronic conditions such as heart failure, showing that patients with limited baseline health knowledge might benefit from ChatGPT's clear, step-by-step instructions.<sup>15</sup> In a study evaluating ChatGPT's responses to atrial fibrillation queries, the chatbot delivered at least partially correct information 98.4% of the time,<sup>16</sup> suggesting that it can be a valuable supplement to care—though human verification remains paramount.

### Performance on specialty examinations within the field of medicine

LLMs have demonstrated a range of scores on medical exams. GPT-3.5 fell below the passing threshold on general practitioner examinations.<sup>17</sup> By contrast, GPT-4 either nearly reached or exceeded the passing level on the European Exam in Core Cardiology (EECC) for text-only questions.<sup>2,18</sup> However, once image-based questions were included (eg, ECG interpretation), ChatGPT scored below the passing level.<sup>2</sup>

For echocardiography boards, GPT-4 has shown limited success, with a correct rate hovering around 50%-55%—underscoring the dependence on text-based prompts rather



**Figure 1.** An overview of a large language model (LLM) framework. info, information.

than use of direct image interpretation.<sup>19</sup> Although promising, these findings reveal that LLMs still lack a complete skill set for comprehensive clinical evaluations in cardiology.

### Role in imaging

Cardiovascular imaging is essential for diagnosis, prognosis, and management. At present, standard LLMs such as GPT-4 cannot “see” images. Nonetheless, some authors have tested GPT-4’s ability to interpret ECG findings (Fig. 3) when all relevant waveforms and intervals are described painstakingly in text form.<sup>20</sup> In these contrived scenarios, GPT-4 sometimes matched or exceeded the performance of human cardiologists in answering multiple-choice questions. Yet, the model remains reliant on the human user to identify and accurately encode the ECG details.<sup>21</sup>

Ongoing research seeks to integrate LLMs with computer vision modules, such as ChatGPT Vision, to bridge between text-based reasoning and raw image interpretation. If successful, LLMs could eventually autonomously interpret echocardiograms or nuclear imaging studies; however, the technology has not yet reached that point.<sup>22</sup>

### Patient communication and social-media forums

Increasingly, patients are seeking medical advice through online forums or social media.<sup>23</sup> One study found that an AI-driven chatbot sometimes provided more empathetic responses than physicians responding on a Reddit forum, potentially impacting patient satisfaction.<sup>24</sup> However, the probability of the chatbot providing misinformation or incomplete guidance remains high if no clinician moderates the interaction. In cardiology, an area in which nuanced risk–benefit discussions are common, careful design and oversight of such chatbot platforms are crucial.

### Cohort phenotyping and risk prediction

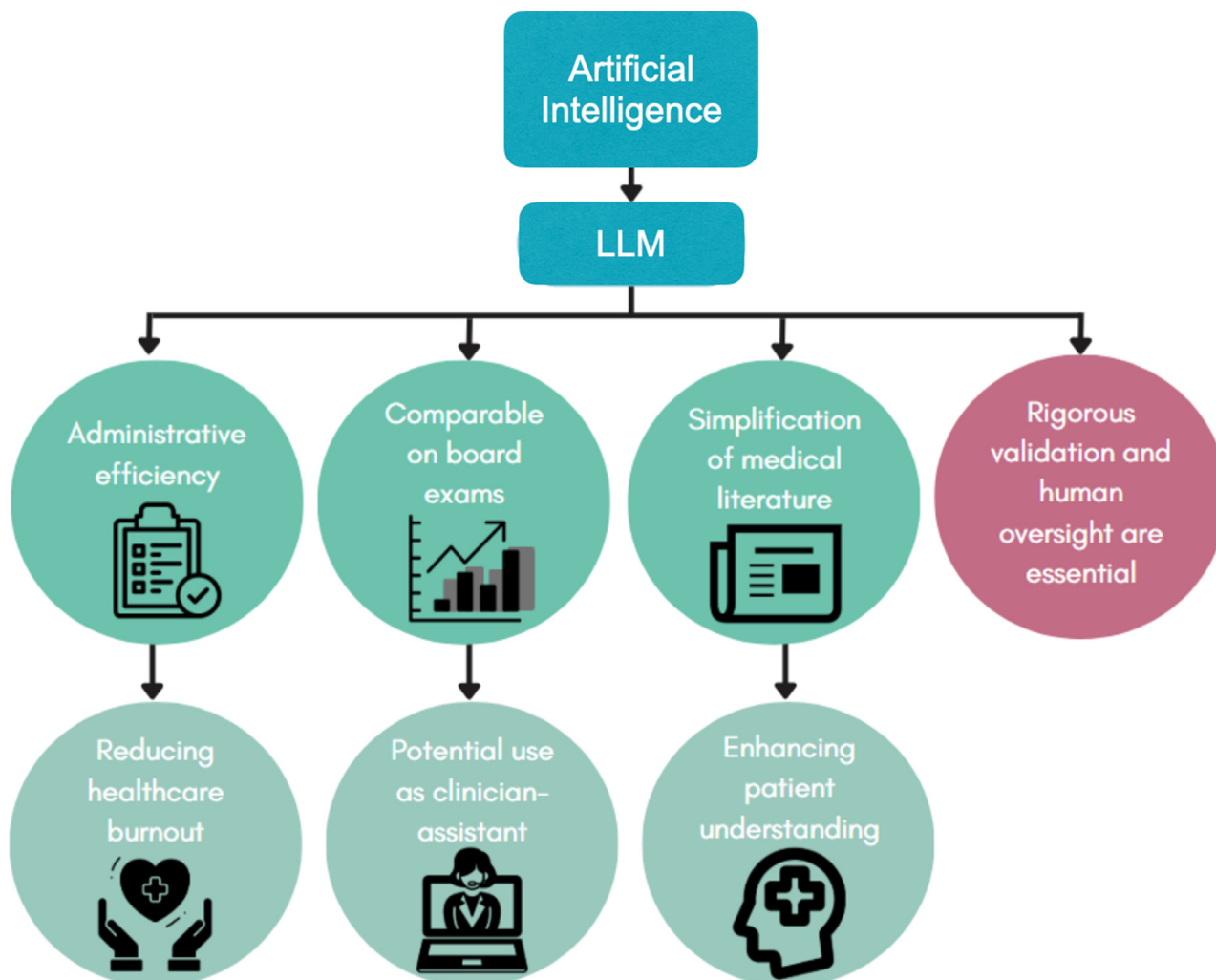
Another frontier for LLMs is patient cohort phenotyping—that is, identifying subsets of patients who are at higher risk for adverse outcomes by analyzing unstructured text from electronic health records. Preliminary studies suggest that LLMs can detect patterns associated with heart failure or myocardial infarction, offering the potential for earlier interventions.<sup>25</sup> With further development, LLMs might deliver predictive analytics for readmissions or complications, although real-world trials addressing this possibility remain sparse.<sup>26</sup>

### Reducing administrative burden

Paperwork and prior-authorization letters occupy a significant portion of clinicians’ time, often contributing to burnout.<sup>27</sup> LLM-driven tools can automatically draft appeals for insurance denials, request letters for coverage of novel therapies, or discharge summaries in plain English.<sup>28</sup> Doximity, for example, integrated ChatGPT in a beta-tool that streamlines digital faxing.<sup>29</sup> This development potentially can free cardiovascular specialists to focus on direct patient care while minimizing clerical tasks.

### ChatGPT vision for cardiologists

In September 2023, OpenAI introduced an update—“ChatGPT can now see, hear, and speak”—expanding ChatGPT’s multimodal capabilities to analyze certain image-based inputs.<sup>30</sup> Although the new module, ChatGPT Vision, allows the model to interpret and describe basic visual content, its performance in specialized clinical imaging—such as ECG waveforms, echocardiography clips, or nuclear scans—remains untested. Although one might hope that use of ChatGPT Vision would reduce reliance on text-based prompts for describing images, in a cardiology setting, it



**Figure 2.** An overview of applications in healthcare using AI and large language models (LLMs).

poses several unique challenges. No regulatory authority has approved this technology for clinical cardiac imaging analysis, and uploading patient images raises concerns about privacy and identifiable information, demanding robust de-identification and secure data handling. Interpreting subtle waveform findings or dynamic imaging such as echocardiogram loops may exceed the model’s current capacity, and randomized trials or prospective real-world evaluations are needed to establish the clinical utility and safety of model use.

If ChatGPT Vision evolves to handle complex cardiac imaging, it could theoretically assist in triage or flag gross abnormalities to aid clinicians working under time constraints. However, at this juncture, such capabilities remain speculative. As with other areas of LLM deployment, thorough validation and human oversight are crucial to ensure patient safety.

### **Ethical Considerations: Hallucinations, Bias, and Privacy**

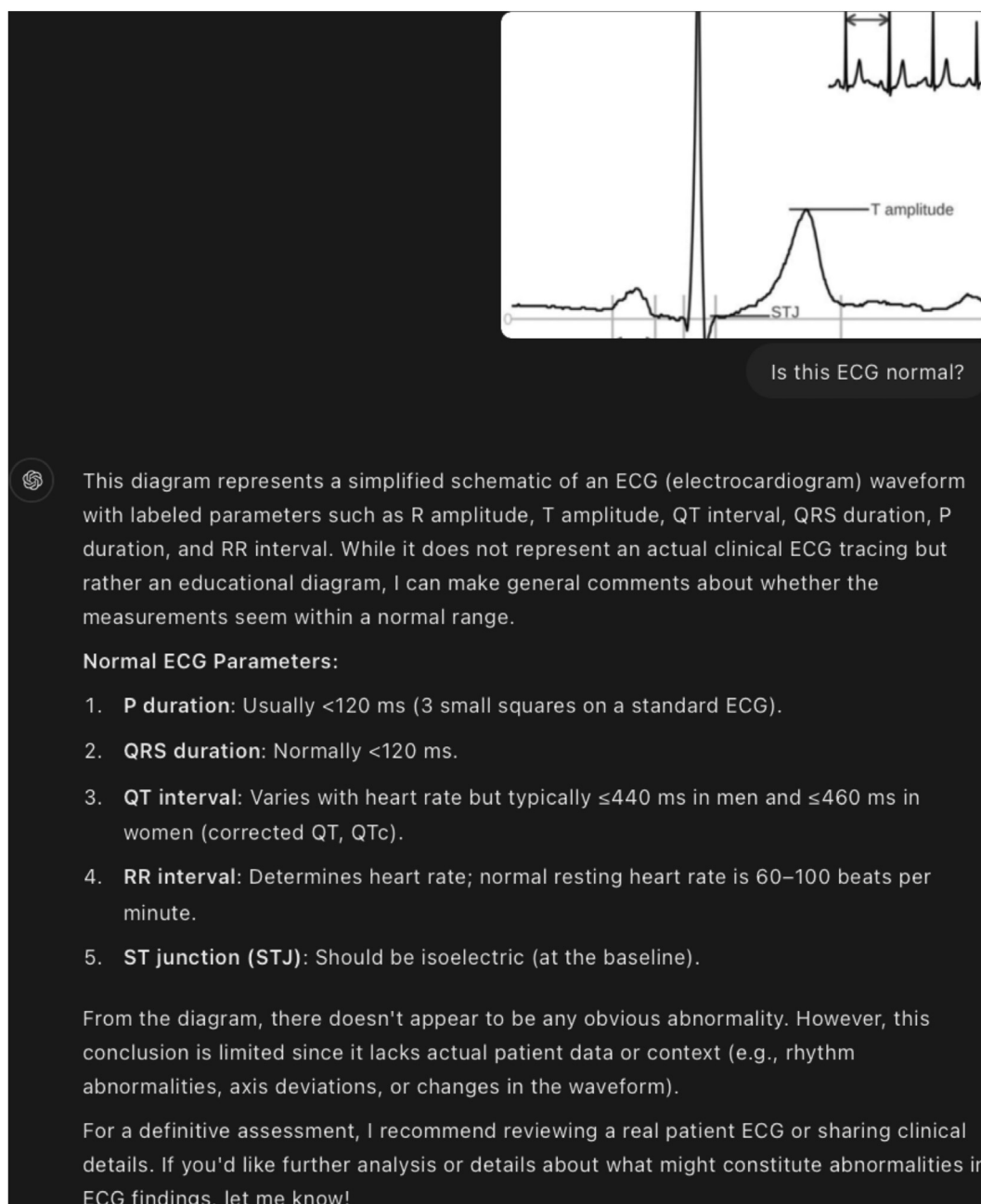
LLMs can produce “hallucinations”—plausible-sounding but entirely incorrect statements or references.<sup>31</sup> This

phenomenon arises because the model is trained to predict the next-most-likely token, not necessarily the most-truthful token. Verification by a knowledgeable human is essential, especially in high-stakes contexts such as cardiology.<sup>32</sup>

Bias in training data also looms large. If the model’s training set underrepresents certain ethnicities or includes biased text, it may produce systematically skewed results.<sup>3</sup> Privacy concerns likewise demand attention. Uploading patient-identifiable data to third-party servers can violate the Health Insurance Portability and Accountability Act (HIPAA) or other local privacy laws unless robust means of de-identification and gathering consent are in place.

Moreover, advanced cardiac care can involve ethically complex decisions—for instance, the deactivation of implantable cardioverter-defibrillators when patients are near the end-of-life stage.<sup>33,34</sup> Although an LLM can supply educational scripts, genuine empathic communication and shared decision-making remain the physician’s domain.

Regulatory bodies, including the World Health Organization, have begun issuing guidance for AI governance in healthcare, urging robust validation, transparent oversight, and user training.<sup>35</sup> These frameworks are crucial to deploying LLMs responsibly in cardiovascular practice.



**Figure 3.** Electrocardiogram (ECG) interpretation by ChatGPT Vision. STJ, sinotubular junction.

### Future Directions

Next-generation models may incorporate multimodal capabilities, allowing direct analysis of ECG tracings, echocardiograms, coronary angiograms, nuclear imaging, and magnetic resonance imaging scans, thereby broadening their role in cardiology diagnostics. Integrating LLMs with electronic health record systems could facilitate real-time risk prediction, generation of personalized reminders, and advanced cohort phenotyping. Prospective clinical trials, clear regulatory pathways, and continuing refinement of ethical guidelines are essential to ensuring safe implementation of such measures.

Research priorities include the following:

- (i) conducting prospective testing in cardiology—evaluating LLM performance in real-world workflows, such as triage, patient counselling, and medication reconciliation;
- (ii) enhancing vision—language models—developing specialized AI systems that interpret images and generate clinically contextualized text;
- (iii) addressing misinformation—creating internal “fact-checking” layers or medical knowledge graphs that reduce the incidence of “hallucinations”; and
- (iv) conducting implementation science—developing understanding of how LLMs influence clinician workflow, patient satisfaction, and health outcomes, and adapting them accordingly.



## Conclusion

ChatGPT and similar LLMs hold promise for cardiovascular professionals, offering new avenues for patient education, administrative efficiency, and potential support in clinical reasoning—particularly for instances in which text-based solutions suffice. However, current models lack robust image-interpretation skills, can “hallucinate” convincing misinformation, and risk amplifying biases. The recent expansion into ChatGPT Vision highlights the technology’s trajectory toward multimodal analysis, but practical cardiology applications remain unvalidated. Overcoming these limitations requires collaborative efforts by clinicians, AI researchers, and policymakers, to establish rigorous validation, ethical guardrails, and responsible deployment. With application of measured optimism and careful oversight regarding their development, LLMs could evolve into valuable partners in the delivery of high-quality cardiology care.

## Acknowledgements

The authors extend gratitude to Alyssa Quizeo for her assistance with creating Figure 2. The authors extend gratitude to Yvonne Leung, PhD, for giving us permission to edit her flow diagram (Fig. 1) to explain the core elements of an LLM.

## Ethics Statement

The authors confirm that patient consent is not applicable to this article. This study is based on publicly accessible information, and does not contain any personally identifiable information; therefore, no patient consent was required.

## Patient Consent

The authors confirm that patient consent is not applicable to this article because it is a narrative review and does not include identifiable patient details or images.

## Funding Sources

The authors have no funding sources to declare.

## Disclosures

The authors have no conflicts of interest to disclose.

## References

- OpenAI. Introducing ChatGPT. Available at: <https://openai.com/blog/chatgpt>. Accessed October 21, 2024.
- Plummer C, Mathysen D, Lawson C. Does ChatGPT succeed in the European Exam in Core Cardiology? *Eur Heart J Digit Health* 2023;4:362-3.
- Wang C, Liu S, Yang H, et al. Ethical considerations of using ChatGPT in health care. *J Med Internet Res* 2023;25:e48009.
- Martineau K. What is generative AI?. Available at: <https://research.ibm.com/blog/what-is-generative-AI>. Accessed October 21, 2024.
- Verma N. Behind the bot: How ChatGPT works and the models powering its Natural Language Processing. Available at: <https://medium.com/@technilesh/behind-the-bot-how-chatgpt-works-and-the-models-powering-its-natural-language-processing-623e7a284f9b>. Accessed October 21, 2024.
- Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30:5998-6008.
- Collins ER. Transformer vs LSTM: a helpful illustrated guide. Available at: <https://blog.finxtor.com/transformer-vs-lstm/>. Accessed October 21, 2024.
- Medium. GPT-4. Available at: <https://openai.com/gpt-4>. Accessed October 21, 2024.
- A.I. Google. Introducing PaLM 2. Available at: <https://blog.google/technology/ai/google-palm-2-ai-large-language-model/>. Accessed October 21, 2024.
- Meta. Llama 2: Open foundation and fine-tuned chat models. Available at: <https://ai.meta.com/research/publications/llama-2-open-foundation-and-fine-tuned-chat-models/>. Accessed October 21, 2024.
- Shahid R, Shoker M, Chu LM, et al. Impact of low health literacy on patients’ health outcomes: a multicenter cohort study. *BMC Health Serv Res* 2022;22:1148.
- Krittawong C, Virk HUH, Kaplin SL, et al. Assessing the potential of ChatGPT for patient education in cardiac catheterization care. *JACC Cardiovasc Interv* 2023;16:1551-2.
- Behers BJ, Vargas IA, Behers BM, et al. Assessing the readability of patient education materials on cardiac catheterization from artificial intelligence chatbots: an observational cross-sectional study. *Cureus* 2024;16:e63865.
- Vahdat S, Hamzehgardeshi L, Hessam S, et al. Patient involvement in health care decision making: a review. *Iran Red Crescent Med J* 2014;16:e12454.
- Dimitriadis F, Alkagiet S, Tsigkriki L, et al. ChatGPT and patients with heart failure [e-pub ahead of print]. *Angiology* 2024. <https://doi.org/10.1177/00033197241238403>.
- Lee TJ, Campbell DJ, Rao AK, et al. Evaluating ChatGPT responses on atrial fibrillation for patient education. *Cureus* 2024;16:e61680.
- Thirunavukarasu AJ, Hassan R, Mahmood S, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test. *JMIR Med Educ* 2023;9:e46599.
- Skalidis I, Cagnina A, Luangphiphat W, et al. ChatGPT takes on the European Exam in Core Cardiology: an AI success story? *Eur Heart J Digit Health* 2023;4:279-81.
- Kangiszer G, Mahtani AU, Pintea M, et al. Low performance of ChatGPT on echocardiography board review questions. *JACC Cardiovasc Imaging* 2024;17:330-2.
- Günay S, Öztürk A, Özerol H, et al. Comparison of emergency medicine specialist, cardiologist, and ChatGPT in electrocardiography assessment. *Am J Emerg Med* 2024;80:51-60.
- Arnaout R. ChatGPT helped me write this talk title, but can it read an echocardiogram? *J Am Soc Echocardiogr* 2023;36:1021-6.
- Fang C, Wu Y, Fu W, et al. How does ChatGPT-4 perform on non-English national medical licensing examination? An evaluation in Chinese language. *PLOS Digit Health* 2023;2:e0000397.
- Lautrup AD, Hyrup T, Schneider-Kamp A, et al. Heart-to-heart with ChatGPT: the impact of patients consulting AI for cardiovascular health advice. *Open Heart* 2023;10:e002455.

24. Ayers JW, Leas EC, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med* 2023;183:589-96.
25. Cabral S, Restrepo D, Kanjee Z, et al. Clinical reasoning of a generative artificial intelligence model compared with physicians. *JAMA Intern Med* 2024;184:581-3.
26. Guillen-Grima F, Guillen-Aguinaga S, Guillen-Aguinaga L, et al. Evaluating the efficacy of ChatGPT in navigating the Spanish Medical Residency Entrance Examination (MIR). *Clin Pract* 2023;13:1460-87.
27. Patel RS, Bachu R, Adikey A, et al. Factors related to physician burnout and its consequences: a review. *Behav Sci (Basel)* 2018;8:98.
28. Herd P, Moynihan D. Health care administrative burdens: centering patient experiences. *Health Serv Res* 2021;56:751-4.
29. Landi H. Doximity rolls out beta version of ChatGPT tool for docs aiming to streamline administrative paperwork. Available at: <https://www.fiercehealthcare.com/health-tech/doximity-rolls-out-beta-version-chatgpt-tool-docs-aiming-streamline-administrative#:~:text=Doximity%2C%20a%20digital%20platform%20for,and%20appeal%20letters%20to%20insurers>. Accessed October 21, 2024.
30. OpenAI. ChatGPT can now see, hear, and speak. Available at: <https://openai.com/index/chatgpt-can-now-see-hear-and-speak>. Accessed October 21, 2024.
31. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on large language models (LLMs). *NPJ Digit Med* 2024;7:183.
32. Balasubramaniam N, Kauppinen M, Rannisto A, et al. Transparency and explainability of AI systems: From ethical guidelines to requirements. *Inf Softw Technol* 2023;159:107197.
33. Bradley AJ, Marks AD. Clinician attitudes regarding ICD deactivation in DNR/DNI patients. *J Hosp Med* 2017;12:498-502.
34. Russo JE. Deactivation of ICDs at the end of life: a systematic review of clinical practices and provider and patient attitudes. *Am J Nurs* 2011;111:26-35.
35. World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. Available at: <https://www.who.int/publications/i/item/9789240084759>. Accessed October 21, 2024.