

Original Paper

ChatGPT's Performance in Cardiac Arrest and Bradycardia Simulations Using the American Heart Association's Advanced Cardiovascular Life Support Guidelines: Exploratory Study

Cecilia Pham¹, BS, EMT; Romi Govender¹, BSN; Salik Tehami¹, BS; Summer Chavez^{1,2,3}, DO, MPH, MPM; Omolola E Adepoju^{1,2,3}, MPH, PhD; Winston Liaw^{1,2,3}, MD, MPH

¹Tilman J Fertitta Family College of Medicine, University of Houston, Houston, TX, United States

²Humana Integrated Health Sciences Institute, University of Houston, Houston, TX, United States

³Department of Health Systems and Population Health Sciences, Tilman J Fertitta Family College of Medicine, Houston, TX, United States

Corresponding Author:

Cecilia Pham, BS, EMT

Tilman J Fertitta Family College of Medicine

University of Houston

5055 Medical Circle

Houston, TX, 77204

United States

Phone: 1 713 743 7047

Email: cmpham4@uh.edu

Abstract

Background: ChatGPT is the most advanced large language model to date, with prior iterations having passed medical licensing examinations, providing clinical decision support, and improved diagnostics. Although limited, past studies of ChatGPT's performance found that artificial intelligence could pass the American Heart Association's advanced cardiovascular life support (ACLS) examinations with modifications. ChatGPT's accuracy has not been studied in more complex clinical scenarios. As heart disease and cardiac arrest remain leading causes of morbidity and mortality in the United States, finding technologies that help increase adherence to ACLS algorithms, which improves survival outcomes, is critical.

Objective: This study aims to examine the accuracy of ChatGPT in following ACLS guidelines for bradycardia and cardiac arrest.

Methods: We evaluated the accuracy of ChatGPT's responses to 2 simulations based on the 2020 American Heart Association ACLS guidelines with 3 primary outcomes of interest: the mean individual step accuracy, the accuracy score per simulation attempt, and the accuracy score for each algorithm. For each simulation step, ChatGPT was scored for correctness (1 point) or incorrectness (0 points). Each simulation was conducted 20 times.

Results: ChatGPT's median accuracy for each step was 85% (IQR 40%-100%) for cardiac arrest and 30% (IQR 13%-81%) for bradycardia. ChatGPT's median accuracy over 20 simulation attempts for cardiac arrest was 69% (IQR 67%-74%) and for bradycardia was 42% (IQR 33%-50%). We found that ChatGPT's outputs varied despite consistent input, the same actions were persistently missed, repetitive overemphasis hindered guidance, and erroneous medication information was presented.

Conclusions: This study highlights the need for consistent and reliable guidance to prevent potential medical errors and optimize the application of ChatGPT to enhance its reliability and effectiveness in clinical practice.

(*J Med Internet Res* 2024;26:e55037) doi: [10.2196/55037](https://doi.org/10.2196/55037)

KEYWORDS

ChatGPT; artificial intelligence; AI; large language model; LLM; cardiac arrest; bradycardia; simulation; advanced cardiovascular life support; ACLS; bradycardia simulations; America; American; heart association; cardiac; life support; exploratory study; heart; heart attack; clinical decision support; diagnostics; algorithms

Introduction

In March 2023, OpenAI, an artificial intelligence (AI) research laboratory, released GPT-4, an experimental version of ChatGPT, a large language model chatbot [1]. Compared with prior models including its predecessor GPT-3.5, GPT-4 has higher accuracy, greater reasoning capabilities, superior confidence, and more safety features with company data demonstrating improved performance on the Uniform Bar Exam and Biology Olympiad [2,3]. Compared with other types of learning models, deep learning models use multiple processing layers and artificial neural networks to process large amounts of data in order for the AI to learn [4]. In health care, ChatGPT shows promising applications including passing the United States Medical Licensing Examination, individualized health advice, improved diagnostic care, faster pharmacologic discovery, task automation, and clinical decision support [5-10]. However, limited evidence exists regarding ChatGPT's ability to perform in advanced clinical scenarios, including cardiac emergencies.

Currently, the standard of care follows algorithms from the 2020 American Heart Association (AHA) advanced cardiovascular life support (ACLS) guidelines. Learners enrolled in the AHA course take a multiple-choice cognitive examination followed by a series of scenario-based patient cases and skills demonstrations [11]. Although an early study by Fijačko et al [12] concluded ChatGPT was unable to pass the AHA's basic life support and advanced cardiovascular life support (ACLS) examinations, they demonstrated ChatGPT was able to provide relevant and accurate explanations for scenario-based questions without learning from an AHA course. With modification, Zhu et al [13] determined ChatGPT was able to pass the AHA's examinations after changing the research protocol in which 3 responses were generated for each question instead of 1, and assessments were turned into open-ended questions. While rudimentary tests find ChatGPT can quickly give well-tailored answers to questions and scenarios regarding cardiopulmonary resuscitation (CPR) based on published guidelines, this has not been rigorously applied to more complex clinical scenarios requiring higher levels of expertise and training [8].

Even with advances in modern health care, heart disease has remained the leading cause of death for several decades [14]. Approximately 436,000 Americans die from cardiac arrest each year, making it a significant public health issue [15]. Past efforts at reducing cardiac arrests include establishing cardiac arrest registries, increasing public awareness and educational efforts, improving the quality of resuscitation care, strengthening emergency medical services, quality improvement programs, and implementing accreditation standards [16]. ACLS protocols have been well-proven to improve cardiac arrest outcomes, with deviations from the guidelines associated with poorer survival and neurologic outcomes [17-22]. Thus, identifying technologies that can improve adherence to ACLS protocols and in turn, cardiac arrest outcomes, is critical. Cognitive aids, such as pocket cards, mobile apps, and augmented reality glasses, have been shown to increase adherence to ACLS guidelines and the quality of resuscitation in adult simulations for in-hospital cardiac arrests [23-28]. ChatGPT could prove to be a valuable

resource by providing real-time decision support in resource-limited settings or augmenting clinical decision-making. Our study aims to analyze the accuracy of ChatGPT's performance in following the AHA's ACLS algorithms in 2 cardiac event simulations—cardiac arrest and bradycardia.

Methods

We evaluated the accuracy of ChatGPT's responses to 2 simulations based on the 2020 AHA ACLS guidelines with 3 primary outcomes of interest as follows: the mean individual step accuracy, the accuracy score per simulation attempt, and the accuracy score for each algorithm.

Ethical Considerations

In accordance with 45 CFR 46, Subpart A, also known as the Common Rule, our research project involving educational tests, such as ACLS simulations, falls under the exemption outlined in 46.104(d)(1). This exemption applies because the interactions in our study are limited to educational tests and do not involve identifiable human subjects. Because of this, no ethics board review was required for this research.

Initial Observations of ChatGPT's Responses and Adherence to AHA Guidelines

The research team underwent an initial testing phase involving several procedures to observe ChatGPT's responses before developing the scripts used for testing. Initially, two types of trials were proposed: (1) assessing ChatGPT's ability to accurately identify rhythm strips, and (2) evaluating ChatGPT's capability to execute the algorithm provided by the AHA. However, from January to May 2023, ChatGPT lacked the ability to recognize images, including rhythm strips or electrocardiograms (ECGs). Due to this limitation, specific ECG rhythms were provided to ChatGPT without the need for correct identification. The team then tested ChatGPT's ability to execute basic algorithms, including those for bradycardia, tachycardia with a pulse, and cardiac arrest, based on the 2020 AHA Guidelines. At the beginning of the trials, it was confirmed by ChatGPT that it would refer to the 2020 AHA Guidelines for information. During these initial trials, testers observed significant differences between the responses generated by GPT-3.5 and GPT-4. GPT-4 demonstrated greater adherence to the 2020 AHA Guidelines by being able to interpret the algorithmic pathways more accurately. Additionally, testers noted that prompting ChatGPT led to lengthy outputs, unlike the step-by-step approach used in real-life simulations. To emulate real-life testing conditions in ACLS training, the command "What is the next step? Give me one step at a time" was included. This method also mimics how ChatGPT could be used in the real world. As GPT-4 closely followed the 2020 AHA Guidelines and was able to respond to the step-by-step command, it was selected for the actual simulations.

Development of Scripts for User Input and ChatGPT Output

Following the completion of the initial trials, the study authors with prior ACLS training (SC, WL, and CP) developed scripts

based on the testers' observations, the AHA's algorithms, and clinical experience. SC and WL are both physicians in emergency medicine and family medicine, respectively, while CP holds an EMT license with paramedic-level training. The 2 scripts exclusively focused on testing 2 categories of cardiac arrhythmias: bradycardia and cardiac arrest ([Multimedia Appendices 1 and 2](#)). To validate the scripts, 3 physicians with ACLS and emergency medicine backgrounds assessed them to ensure they accurately depict real-world clinical scenarios and management practices. The study authors developed a scoring guide to measure the accuracy of the responses of ChatGPT to prespecified prompts. The number of attempts given to ChatGPT was matched to the required correct outputs for each group. For instance, if 5 correct outputs were expected, ChatGPT was prompted 5 times, allowing it the same number of opportunities to provide accurate responses. In instances where ChatGPT's outputs diverged from the predefined script, the authors responded by typing "Give me one step at a time" or "Give me another option."

Data Collection

Testing was conducted between May to August 2023 using the public web user chat interface by OpenAI, with upgraded accounts to use GPT-4. All inputs and outputs for 1 simulation are saved within a single conversation thread, and each simulation was conducted in a separate thread. Three testers underwent training to adhere to the scripts, ensuring standardized procedures for the simulations. Each scenario was tested 20 times to evaluate the variability of ChatGPT's responses and its ability to provide correct outputs. For each simulation step, ChatGPT was scored for correctness (1 point) or incorrectness (0 points), and these scores were recorded in an Excel (Microsoft Corp) spreadsheet. Partially correct responses were considered incorrect. Within each scoring section, qualitative comments regarding the incorrect responses were also recorded. Additionally, instances where the prompt "Give me one step at a time" requiring repeated inputs were recorded. The full transcripts for each simulation were collated into separate

document files for each simulation attempt. The authors of the scripts independently reviewed the recorded accuracy scores in the spreadsheets and the transcripts saved in the document files to ensure the validity of the results and to maintain accurate record-keeping.

Data Analysis

Using the data recorded in the Excel spreadsheets, accuracy scores for individual steps, individual simulation attempts, and overall simulation attempts for each algorithm were calculated. The overall score per simulation attempt was calculated by summing the correct responses and dividing them by the number of steps (12 for bradycardia and 39 for cardiac arrest). For each step in the simulation, we similarly summed the correct responses and divided them by the number of simulations ($n=20$). Mean overall simulation accuracies were determined by averaging 20 simulation scores. The median accuracies for each algorithm were also calculated. Qualitative comments recorded for each simulation were also analyzed to identify recurring themes to describe ChatGPT's performance.

Results

Overview

The individual step accuracy per simulation attempt for the cardiac arrest and bradycardia algorithms are reported in [Tables 1 and 2](#), respectively. ChatGPT's median accuracy for each step was 85% (IQR 40%-100%) for cardiac arrest and 30% (IQR 13%-81%) for bradycardia. The accuracy scores per simulation attempt for each algorithm are described in [Table 3](#). ChatGPT's median accuracy for over 20 simulation attempts for cardiac arrest was 69% (IQR 67%-74%) and for bradycardia was 42% (IQR 33%-50%). Four key findings were identified after analyzing the results from the 2 sets of simulations: (1) ChatGPT's outputs varied despite consistent input, (2) the same actions were persistently missed, (3) repetitive overemphasis hindered guidance, and (4) erroneous medication information was presented.

Table 1. Individual step accuracy using ChatGPT in cardiac arrest advanced cardiovascular life support simulations (N=20).

Correct simulation output	Accuracy, n (%) ^a
Part 1: Initiation of CPR^b	
Assess for responsiveness	20 (100)
Assess for a pulse	20 (100)
Activate an emergency response system	19 (95)
Start CPR	20 (100)
Apply the defibrillator	20 (100)
Obtain intravenous or intraosseous access	2 (10)
Part 2: First dose of epinephrine	
Assess for responsiveness	0 (0)
Assess for a pulse	4 (20)
Continue CPR	17 (85)
Give 1 mg of epinephrine	19 (95)
Consider advanced airway management	13 (65)
“Continue CPR” after completing Part 2	16 (80)
Part 3: First rhythm check	
Perform a rhythm check	20 (100)
Consider reversible causes	8 (40)
Continue CPR	20 (100)
Part 4: Second dose of epinephrine	
Perform a rhythm check	12 (60)
Immediately defibrillate	11 (55)
Give 1 mg of epinephrine	15 (75)
Consider reversible causes	4 (20)
“Continue CPR” after completing Part 4	20 (100)
Part 5: Second defibrillation	
Perform a rhythm check	18 (90)
Immediately defibrillate	19 (95)
“Continue CPR” after completing Part 5	20 (100)
Part 6: First dose of alternative pharmacologic agents	
Give 1 mg of epinephrine	6 (30)
Consider giving amiodarone or lidocaine	9 (45)
Consider reversible causes	0 (0)
Perform a rhythm check	20 (100)
Immediately defibrillate	20 (100)
“Continue CPR” after completing Part 6	16 (80)
Part 7: Additional pharmacologic agents	
Perform a rhythm check	19 (95)
Immediately defibrillate	18 (90)
Consider reversible causes	2 (10)
Give a second dose of amiodarone	6 (30)
“Continue CPR” after completing Part 7	19 (95)

Correct simulation output	Accuracy, n (%) ^a
Part 8: End of simulation	
Give 1 mg of epinephrine	14 (70)
Consider giving lidocaine	0 (0)
Consider reversible causes	11 (55)
Perform a rhythm check	18 (90)
“Continue CPR” after completing Part 8	20 (100)

^aMedian accuracy is 85% (IQR 40%-100%); mean accuracy is 69%.

^bCPR: cardiopulmonary resuscitation.

Table 2. Individual step accuracy using ChatGPT in bradycardia advanced cardiovascular life support simulations (N=20). Mean individual step accuracy using ChatGPT in bradycardia advanced cardiovascular life support simulations (N=20).

Correct simulation output	Accuracy, n (%) ^a
Part 1: Address ABC's of resuscitation	
Maintain the patient's airway; assists breathing as necessary	16 (80)
Apply a cardiac monitor to identify rhythm	3 (15)
Monitor blood pressure, pulse oximetry, and other vital signs	9 (45)
Obtain intravenous access	7 (35)
Obtain 12-lead ECG ^b	0 (0)
Give 1 mg of atropine	0 (0)
Part 2: Address hypoxia	
Give oxygen	20 (100)
Part 3: Second dose of atropine	
Give an additional dose of atropine	19 (95)
Part 4: Alternative pharmacologic agents and consideration of pacing	
Give an additional dose of atropine	3 (15)
Consider transcutaneous pacing	17 (85)
Give dopamine or epinephrine	1 (5)
Part 5: Seek expert consultation	
Seek a consultation	5 (25)

^aMedian accuracy is 30% (IQR 13%-81.3%); mean accuracy is 42%.

^bECG: electrocardiogram.

Table 3. Accuracy score per simulation attempt for ChatGPT in cardiac arrest (N=39) and bradycardia (N=12) advanced cardiovascular life support simulations.

Simulation number	Accuracy of cardiac arrest, n (%) ^a	Accuracy of bradycardia, n (%) ^b
1	24 (62)	7 (58)
2	26 (67)	5 (42)
3	24 (62)	6 (50)
4	29 (74)	6 (50)
5	27 (69)	5 (42)
6	27 (69)	7 (58)
7	21 (54)	4 (33)
8	29 (74)	4 (33)
9	26 (67)	7 (58)
10	27 (69)	6 (50)
11	26 (67)	5 (42)
12	26 (67)	5 (42)
13	22 (56)	5 (42)
14	26 (67)	5 (42)
15	27 (69)	3 (25)
16	29 (74)	7 (58)
17	29 (74)	3 (25)
18	30 (77)	5 (42)
19	30 (77)	3 (25)
20	30 (77)	4 (33)
Average accuracy	535/780 (69)	102/240 (43)

^aMedian accuracy of cardiac arrest is 69% (IQR 67%-74%).

^bMedian accuracy of bradycardia is 42% (IQR 33%-50%).

ChatGPT's Outputs Varied

ChatGPT gave varying responses to identical inputs, sometimes providing unexpected responses. ChatGPT generated multistep instructions, prompting the request "give me one step at a time." It would sometimes provide entirely different steps that were not in its original multistep output. There were, on average, 0.5 instances per simulation of repetitive loops of "check heart rhythm" and "resume CPR." Even when "one step at a time" was requested, ChatGPT was unable to consistently execute algorithms step by step, often generating multistep instructions. ChatGPT was instructed to give "one step at a time" 3.6 times per simulation in the cardiac arrest case and 1.5 times per simulation in the bradycardia case. Other unexpected outputs include 0.5 instances per simulation of oxygen and medication delivery instructions preceding necessary steps such as applying a monitor to the patient or starting an intravenous (IV) line. ChatGPT provided additional instructions that were not part of the ACLS algorithms, including 0.3 instances per simulation of "start an isoproterenol infusion."

Some Actions Were Persistently Missed

Throughout the simulations, there were several notable omissions to ChatGPT's output. ChatGPT frequently skipped

the instruction to establish IV or intraosseous (IO) access, before directing IV or IO medication administration at 0% accuracy in cardiac arrest and 35% (n=7) accuracy in bradycardia. ChatGPT also struggled to incorporate additional dosages of medications in its responses. For epinephrine administration, ChatGPT was accurate 95% (n=19) of the time regarding the first dose, but its accuracy decreased to 75% (n=15) and 70% (n=14) for the second and third doses, respectively. In the bradycardia simulations, ChatGPT consistently recommended a second dose of atropine at 95% (n=19) accuracy but rarely advised a third dose at 15% (n=3) accuracy. Furthermore, ChatGPT was never prompted to obtain a 12-lead ECG with 0% accuracy. ChatGPT did not provide any instructions for checking vital signs and only suggested pulse checks for the patient with bradycardia, omitting other vital signs such as blood pressure and oxygen saturation.

Repetitive Overemphasis Hindered Guidance

ChatGPT consistently overemphasized specific steps to the detriment of other important interventions. During the cardiac arrest simulations, ChatGPT became fixed in repetitive cycles when given the prompt, "CPR resumed. It's been two minutes. What is the next immediate step?" ChatGPT would then incessantly advise to "check the heart rhythm" and "resume

CPR” on average 4 times per simulation without providing any other instructions such as considering advanced airway management at 65% (n=13) accuracy and exploring reversible causes with 25% (n=5) accuracy. Similarly, in the bradycardia simulation, ChatGPT emphasized transcutaneous and transvenous pacing methods on average 1.8 times per simulation, disregarding other reasonable alternatives for bradycardia management, such as pharmacologic interventions.

Erroneous Medication Information Was Presented

Two medication instructions given in the bradycardia simulations were consistently inaccurate. In all bradycardia simulations, ChatGPT advised administering 0.5 mg of atropine instead of the recommended 1 mg dose in the most recent 2020 AHA ACLS guidelines (0% accuracy). Despite being corrected and acknowledging the accurate dosage as 1 mg, ChatGPT persisted in administering the incorrect dose of atropine in subsequent outputs. In 95% (n=19) of bradycardia simulations, ChatGPT provided incorrect guidance to administer 2-20 µg/kg/minute of dopamine, deviating from the correct range of 5-20 µg/kg/minute. However, in all cardiac arrest simulations, all medication dosages were correctly identified by ChatGPT.

Discussion

Principal Findings

This study sought to examine the accuracy of ChatGPT’s performance in adult cardiac event simulations for cardiac arrest and bradycardia. Although the overall accuracy of ChatGPT was lower than the standard passing threshold in academics of 70%, the demonstrated accuracy of ChatGPT was comparable to one other study of complex cardiac clinical vignettes which had 50% accuracy (50/100). Compared with experts, ChatGPT gave inaccurate or incomplete responses [29]. Even with these caveats, this proof-of-concept study illustrates the potential and perils of using ChatGPT for real-time decision support in clinical settings.

Our study adds to the small but growing body of literature analyzing the potential clinical role of ChatGPT and other AI technologies, with a special focus on cardiac emergencies and complex decision-making. While this field may be in its infancy, ChatGPT builds upon prior generations of more low-tech cognitive aids such as apps or paper cards by offering faster responses, tailored answers, and the ability to work through far more complicated scenarios. In several steps of the simulations, ChatGPT’s accuracy was as high as 100% (n=20), an impressive benchmark.

In practice settings with limited resources, ChatGPT, AI, and related technologies may still help fill a crucial gap by acting as a more advanced cognitive aid compared with previously studied solutions. They can also be used in medical education, quality initiatives, and simulation training exercises. However, the findings from this study highlight several key issues that need to be addressed before implementing this decision aid in patient care.

In our study, ChatGPT exhibited a wide range of clinical accuracy, which can be due in part to outdated or unreliable training data. If AI is used to augment clinical decision-making,

physicians must ensure the appropriateness and validity of the training data. Without this level of accountability, health care professionals could accept hallucinations, or AI responses that are incorrect or misleading. While subject matter experts may detect these mistakes, those with less training and experience may fail to recognize these inconsistencies, which can lead to missing steps or delivering inaccurate medication doses. Additionally, if the training data contains inaccuracies or reflects certain biases, ChatGPT may replicate those errors or biases in its responses. It is also unclear how frequently ChatGPT updates or how it decides to include new or controversial scientific findings in its model. This is especially salient in cardiac emergencies, where quick and accurate decision-making is crucial for patient survival.

Our findings suggest that ChatGPT has difficulty learning from past scenarios. For example, ChatGPT would acknowledge the correct dosage if it initially gave an incorrect one, but persistently administered the same incorrect doses across subsequent trials. Real-life ACLS scenarios involve complex medical situations with multiple possible interventions depending on the patient’s status. Patients with cardiac arrest secondary to hyperkalemia or pulmonary embolism are managed differently than the simpler cases created for this simulation. Achieving higher levels of accuracy would either require the ability for ChatGPT to learn or a different AI model altogether.

Some experts recommend asking probing questions to AI as a possible solution to verify responses from ChatGPT. In clinical scenarios where time directly correlates with myocardial health and patient outcomes, this may be unrealistic. Continuing to prompt the AI or slightly modifying responses to see how responses change is also not practical. In our study, ChatGPT was prompted dozens of times for responses. During an actual episode of cardiac arrest, where intervals of CPR are done in 2 minutes, spending most of the time repeatedly prompting ChatGPT for an accurate response is probably not time well spent during the resuscitation.

Several limitations should be considered when interpreting these results. First, our study focused only on 2 ACLS scenarios: bradycardia and cardiac arrest. Because this is a proof-of-concept study, the authors decided to test just 2 scenarios. In the overall ACLS framework, there are many other emergencies covered, including the management of tachyarrhythmias, opioid overdose, and stroke. Given that ChatGPT’s performance could vary in these alternative scenarios, broadening the scope of simulations might provide a more thorough insight into ChatGPT’s capabilities by evaluating its performance across a broader range of ACLS conditions. Our study analyzed the accuracy of ChatGPT only, making it difficult to generalize the results to other large language models. During the study period, ChatGPT could not recognize images, requiring the study authors to specify the cardiac rhythm in the model, potentially affecting the accuracy. There are growing concerns that the quality of ChatGPT has decreased over time as it has become more widely used by the public, which could also impact the results [30]. Additionally, given the potential evolution of ChatGPT’s capabilities over time, the reliability and validity of the study’s results may be affected. Exploring how AI models can update their knowledge

could provide insights into their long-term utility in real-world medical emergency settings.

Conclusions

Bridging the gap between AI and human language is a multifaceted challenge, balancing advancements, data quality, and human oversight to maximize ChatGPT's benefits and minimize risks in health care. This study highlights the need for consistent and reliable guidance to prevent potential medical errors and optimize the application of ChatGPT to enhance its reliability and effectiveness in clinical practice. To enhance

ChatGPT's real-world effectiveness, significant improvements are needed, particularly in accuracy and consistency across diverse medical situations. While ChatGPT holds promise as a decision support tool that can provide structured clinical guidance, it should complement, not replace, qualified health care professionals. Future research also should aim to address these limitations and further investigate the challenges of AI in health care to ensure its safe and effective use. Additionally, future studies comparing ChatGPT's performance with other AI models or with human performance could offer valuable insights.

Data Availability

The data sets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

ChatGPT ACLS Testing Algorithm for Bradycardia.

[\[PNG File, 175 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

ChatGPT ACLS Testing Algorithm for Cardiac Arrest.

[\[PNG File, 311 KB-Multimedia Appendix 2\]](#)

References

1. Introducing ChatGPT. OpenAI. URL: <https://openai.com/blog/chatgpt> [accessed 2023-08-07]
2. GPT-4. AI O. 2023. URL: <https://openai.com/gpt-4> [accessed 2023-10-20]
3. Brin D, Sorin V, Vaid A, Soroush A, Glicksberg BS, Charney AW, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. *Sci Rep*. 2023;13(1):16492. [FREE Full text] [doi: [10.1038/s41598-023-43436-9](https://doi.org/10.1038/s41598-023-43436-9)] [Medline: [37779171](https://pubmed.ncbi.nlm.nih.gov/37779171/)]
4. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. [doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)] [Medline: [26017442](https://pubmed.ncbi.nlm.nih.gov/26017442/)]
5. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198. [FREE Full text] [doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)] [Medline: [36812645](https://pubmed.ncbi.nlm.nih.gov/36812645/)]
6. Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. *BenchCouncil Trans Benchmarks, Stand Eval*. 2023;3(1):100105. [FREE Full text] [doi: [10.1016/j.tbench.2023.100105](https://doi.org/10.1016/j.tbench.2023.100105)]
7. Liu S, Wright AP, Patterson BL, Wanderer JP, Turer RW, Nelson SD, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *J Am Med Inform Assoc*. 2023;30(7):1237-1245. [FREE Full text] [doi: [10.1093/jamia/ocad072](https://doi.org/10.1093/jamia/ocad072)] [Medline: [37087108](https://pubmed.ncbi.nlm.nih.gov/37087108/)]
8. Ahn C. Exploring ChatGPT for information of cardiopulmonary resuscitation. *Resuscitation*. 2023;185:109729. [FREE Full text] [doi: [10.1016/j.resuscitation.2023.109729](https://doi.org/10.1016/j.resuscitation.2023.109729)] [Medline: [36773836](https://pubmed.ncbi.nlm.nih.gov/36773836/)]
9. Waqas. AI in healthcare: ChatGPT helps boy get diagnosis after doctors fail. *HACK Read*. 2023. URL: <https://www.hackread.com/artificial-intelligence-healthcare-chatgpt-boy-diagnosis/> [accessed 2024-04-04]
10. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? the implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. [FREE Full text] [doi: [10.2196/45312](https://doi.org/10.2196/45312)] [Medline: [36753318](https://pubmed.ncbi.nlm.nih.gov/36753318/)]
11. 2020 HeartCode® courses FAQs. *cpr.heart.org*. 2020. URL: https://cpr.heart.org/-/media/cpr-files/course-faqs/2020-course-faqs/2020-heartcode-courses-faqs_102120.pdf?la=en [accessed 2023-09-11]
12. Fijačko N, Gosak L, Štiglic G, Picard CT, Douma MJ. Can ChatGPT pass the life support exams without entering the American heart association course? *Resuscitation*. 2023;185:109732. [FREE Full text] [doi: [10.1016/j.resuscitation.2023.109732](https://doi.org/10.1016/j.resuscitation.2023.109732)] [Medline: [36775020](https://pubmed.ncbi.nlm.nih.gov/36775020/)]
13. Zhu L, Mou W, Yang T, Chen R. ChatGPT can pass the AHA exams: open-ended questions outperform multiple-choice format. *Resuscitation*. 2023;188:109783. [FREE Full text] [doi: [10.1016/j.resuscitation.2023.109783](https://doi.org/10.1016/j.resuscitation.2023.109783)] [Medline: [37349064](https://pubmed.ncbi.nlm.nih.gov/37349064/)]

14. Heart disease deaths—health, United States. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/nchs/hus/topics/heart-disease-deaths.htm> [accessed 2023-08-07]
15. CPR facts and stats. cpr.heart.org. URL: <https://cpr.heart.org/en/resources/cpr-facts-and-stats> [accessed 2024-04-02]
16. Committee on the Treatment of Cardiac Arrest: Current Status and Future Directions, Board on Health Sciences Policy, Institute of Medicine. Recommendations and key opportunities. In: Graham R, McCoy MA, Schultz AM, editors. Strategies to Improve Cardiac Arrest Survival: A Time to Act. vol 7. Washington (DC). National Academies Press (US); 2015.
17. History of CPR. Association AH. URL: <https://cpr.heart.org/en/resources/history-of-cpr> [accessed 2023-10-20]
18. Crowley CP, Saliccioli JD, Kim EY. The association between ACLS guideline deviations and outcomes from in-hospital cardiac arrest. *Resuscitation*. 2020;153:65-70. [FREE Full text] [doi: [10.1016/j.resuscitation.2020.05.042](https://doi.org/10.1016/j.resuscitation.2020.05.042)] [Medline: [32502576](https://pubmed.ncbi.nlm.nih.gov/32502576/)]
19. Sanders AB, Berg RA, Bures M, Genova RT, Kern KB, Ewy GA. The efficacy of an ACLS training program for resuscitation from cardiac arrest in a rural community. *Ann Emerg Med*. 1994;23(1):56-59. [doi: [10.1016/s0196-0644\(94\)70009-5](https://doi.org/10.1016/s0196-0644(94)70009-5)] [Medline: [8273960](https://pubmed.ncbi.nlm.nih.gov/8273960/)]
20. Sodhi K, Singla MK, Shrivastava A. Impact of advanced cardiac life support training program on the outcome of cardiopulmonary resuscitation in a tertiary care hospital. *Indian J Crit Care Med*. 2011;15(4):209-212. [FREE Full text] [doi: [10.4103/0972-5229.92070](https://doi.org/10.4103/0972-5229.92070)] [Medline: [22346031](https://pubmed.ncbi.nlm.nih.gov/22346031/)]
21. Honarmand K, Mephram C, Ainsworth C, Khalid Z. Adherence to Advanced Cardiovascular Life Support (ACLS) guidelines during in-hospital cardiac arrest is associated with improved outcomes. *Resuscitation*. 2018;129:76-81. [doi: [10.1016/j.resuscitation.2018.06.005](https://doi.org/10.1016/j.resuscitation.2018.06.005)] [Medline: [29885353](https://pubmed.ncbi.nlm.nih.gov/29885353/)]
22. Hunt EA, Vera K, Diener-West M, Haggerty JA, Nelson KL, Shaffner DH, et al. Delays and errors in cardiopulmonary resuscitation and defibrillation by pediatric residents during simulated cardiopulmonary arrests. *Resuscitation*. 2009;80(7):819-825. [doi: [10.1016/j.resuscitation.2009.03.020](https://doi.org/10.1016/j.resuscitation.2009.03.020)] [Medline: [19423210](https://pubmed.ncbi.nlm.nih.gov/19423210/)]
23. Corazza F, Fiorese E, Arpone M, Tardini G, Frigo AC, Cheng A, et al. The impact of cognitive aids on resuscitation performance in in-hospital cardiac arrest scenarios: a systematic review and meta-analysis. *Intern Emerg Med*. 2022;17(7):2143-2158. [FREE Full text] [doi: [10.1007/s11739-022-03041-6](https://doi.org/10.1007/s11739-022-03041-6)] [Medline: [36031672](https://pubmed.ncbi.nlm.nih.gov/36031672/)]
24. Corazza F, Arpone M, Tardini G, Stritoni V, Mormando G, Graziano A, et al. Effectiveness of a novel tablet application in reducing guideline deviations during pediatric cardiac arrest: a randomized clinical trial. *JAMA Netw Open*. 2023;6(8):e2327272. [FREE Full text] [doi: [10.1001/jamanetworkopen.2023.27272](https://doi.org/10.1001/jamanetworkopen.2023.27272)] [Medline: [37535352](https://pubmed.ncbi.nlm.nih.gov/37535352/)]
25. Siebert JN, Ehrler F, Gervais A, Haddad K, Lacroix L, Schrurs P, et al. Adherence to AHA guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: a randomized controlled trial. *J Med Internet Res*. 2017;19(5):e183. [FREE Full text] [doi: [10.2196/jmir.7379](https://doi.org/10.2196/jmir.7379)] [Medline: [28554878](https://pubmed.ncbi.nlm.nih.gov/28554878/)]
26. Siebert JN, Lacroix L, Cantais A, Manzano S, Ehrler F. The impact of a tablet app on adherence to American Heart Association guidelines during simulated pediatric cardiopulmonary resuscitation: randomized controlled trial. *J Med Internet Res*. 2020;22(5):e17792. [FREE Full text] [doi: [10.2196/17792](https://doi.org/10.2196/17792)] [Medline: [32292179](https://pubmed.ncbi.nlm.nih.gov/32292179/)]
27. Hejjaji V, Malik AO, Peri-Okonny PA, Thomas M, Tang Y, Wooldridge D, et al. Mobile app to improve house officers' adherence to advanced cardiac life support guidelines: quality improvement study. *JMIR Mhealth Uhealth*. 2020;8(5):e15762. [FREE Full text] [doi: [10.2196/15762](https://doi.org/10.2196/15762)] [Medline: [32427115](https://pubmed.ncbi.nlm.nih.gov/32427115/)]
28. Brophy SL, McCue MR, Reel RM, Jones TD, Dias RD. The impact of a smartphone-based cognitive aid on clinical performance during cardiac arrest simulations: a randomized controlled trial. *AEM Educ Train*. 2023;7(3):e10880. [FREE Full text] [doi: [10.1002/aet2.10880](https://doi.org/10.1002/aet2.10880)] [Medline: [37292063](https://pubmed.ncbi.nlm.nih.gov/37292063/)]
29. Harskamp RE, De Clercq L. Performance of ChatGPT as an AI-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). medRxiv Preprint posted online on March 26, 2023. 2023.:1-17. [FREE Full text] [doi: [10.1101/2023.03.25.23285475](https://doi.org/10.1101/2023.03.25.23285475)]
30. Chen L, Zaharia M, Zou J. How is ChatGPT's behavior changing over time? ArXiv Preprint posted online on 31 Oct 2023. 2023. [FREE Full text] [doi: [10.5260/chara.21.2.8](https://doi.org/10.5260/chara.21.2.8)]

Abbreviations

- ACLS:** advanced cardiovascular life support
- AHA:** American Heart Association
- AI:** artificial intelligence
- CPR:** cardiopulmonary resuscitation
- ECG:** electrocardiogram
- IO:** intraosseous
- IV:** intravenous

Edited by Q Jin; submitted 30.11.23; peer-reviewed by D Wright, TAR Sure, L Zhu, S Kommireddy; comments to author 08.02.24; revised version received 22.02.24; accepted 10.03.24; published 22.04.24

Please cite as:

Pham C, Govender R, Tehami S, Chavez S, Adepoju OE, Liaw W

ChatGPT's Performance in Cardiac Arrest and Bradycardia Simulations Using the American Heart Association's Advanced Cardiovascular Life Support Guidelines: Exploratory Study

J Med Internet Res 2024;26:e55037

URL: <https://www.jmir.org/2024/1/e55037>

doi: [10.2196/55037](https://doi.org/10.2196/55037)

PMID:

©Cecilia Pham, Romi Govender, Salik Tehami, Summer Chavez, Omolola E Adepoju, Winston Liaw. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 22.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.