

Review

Not peer-reviewed version

---

# ChatGPT Applications in Heart Failure: Patient Education, Readability Enhancement, and Clinical Utility

---

[Robert S. Doyle](#) , [Jack Hartnett](#) \* , [Hugo C. Temperley](#) , [Cian Murray](#) , Ross Walsh , Jamie Walsh , [John McCormick](#) , Catherine McGorrian , Katie Murphy , Kenneth McDonald

Posted Date: 16 October 2025

doi: 10.20944/preprints202510.1260.v1

Keywords: ChatGPT; heart failure; artificial intelligence; patient education; systematic review; language models



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

*Systematic Review*

# ChatGPT Applications in Heart Failure: Patient Education, Readability Enhancement, and Clinical Utility

Robert S. Doyle <sup>1</sup>, Jack Hartnett <sup>1,\*</sup>, Hugo C. Temperley <sup>2,3</sup>, Cian Murray <sup>4</sup>, Ross Walsh <sup>5</sup>, Jamie Walsh <sup>5</sup>, John McCormick <sup>6</sup>, Catherine McGorrian <sup>1</sup>, Katie Murphy <sup>1</sup> and Kenneth McDonald <sup>6</sup>

<sup>1</sup> Cardiology Department, Mater Misericordiae University Hospital, Eccles St., D07 R2WY Dublin, Ireland

<sup>2</sup> Radiology Department, St. James's Hospital, James's St., Saint James, D08 NHY1 Dublin, Ireland

<sup>3</sup> Trinity College Dublin, The University of Dublin, College Green, D02 PN40 Dublin, Ireland

<sup>4</sup> Cardiology Department, St. James's Hospital, James's St., Saint James, D08 NHY1 Dublin, Ireland

<sup>5</sup> Cardiothoracic Surgery Department, Mater Misericordiae University Hospital, Eccles St., D07 R2WY Dublin, Ireland

<sup>6</sup> St. Vincent's University Hospital, Elm Park, D04 T6F4 Dublin, Ireland

\* Correspondence: jhartne@tcd.ie

## Abstract

**Background:** Heart failure (HF) affects over 64 million people globally, imposing substantial morbidity, mortality, and economic burdens. Despite advances in guideline-directed therapies, adherence remains suboptimal due to low health literacy and complex regimens. ChatGPT, an advanced large language model by OpenAI, offers conversational capabilities that could enhance HF education, management, and research. This systematic review synthesizes evidence on ChatGPT's applications in HF, evaluating its accuracy in patient education and question-answering, enhancing readability, and clinical documentation/symptom extraction. **Methods:** Following PRISMA guidelines, we searched PubMed, Embase, and Cochrane up to July 2025 using the terms "ChatGPT" and "heart failure." Inclusion: Studies on ChatGPT (3.5 or 4) in HF contexts, such as in education, readability and symptom extraction. Exclusion: Non-HF or non-ChatGPT AI. Data extraction covered design, objectives, methods, outcomes. Thematic synthesis applied. **Results:** From 59 records, 7 observational studies were included. Themes included: patient education/question-answering (n=5), readability enhancement (n=2), clinical documentation/symptom extraction (n=1). Accuracy ranged 78-98%, with high reproducibility; readability improved to 6th-7th grade levels; and symptom extraction achieved up to 95% F1 score, outperforming traditional machine learning baselines. **Conclusions:** ChatGPT shows promise in HF care but requires further randomized validation for outcomes and bias mitigation.

**Keywords:** ChatGPT; heart failure; artificial intelligence; patient education; systematic review; language models

## 1. Introduction

Heart failure (HF) represents an increasingly prevalent clinical condition, impacting more than 6.7 million individuals in the United States alone [1]. Projections indicate a 46% increase in prevalence, with more than 8 million people expected to be affected by 2030 in the United States [2]. Despite advances in HF management, it continues to impose a substantial financial burden on the US healthcare system, with total costs in the US estimated at USD 30.7 billion in 2012 and projected to increase to USD 69.8 billion by 2030 [3]. Guideline-directed medical therapy (GDMT), combined with intensive patient follow-up, have significantly improved HF-related mortality and reduced

readmission rates in recent decades [6,7]. However, contemporary heart failure management is highly labour intensive, both on patients and healthcare providers. Innovative approaches will be required to meet the increasing demand for heart failure care with finite healthcare resources [8].

Artificial intelligence (AI) and machine learning (ML) techniques have been explored as potential tools in this regard [9]. ML models have demonstrated benefits when studied alongside conventional statistics in various fields of cardiovascular medicine [10]. AI algorithms have the potential to improve HF care by supporting clinical decision-making, optimizing treatment allocation, identifying those who benefit most from therapy, predicting adverse outcomes and detecting patients with sub-clinical disease or worsening HF [29–35]. ML models have also been shown to enhance HF diagnosis by analyzing a wide range of data from sources including electrocardiograms, echocardiography, remote monitoring devices and heart sounds [12,13].

Recently, ChatGPT (Generative Pre-trained Transformer), the state-of-the-art conversational model, has attracted worldwide attention for its capability of generating human-like responses to natural language inputs. As an integral part of OpenAI's pre-training transformer models, it currently represents one of the most extensively accessible language models. With the ability to understand and replicate the intricacies and nuances of human language, ChatGPT is rapidly emerging as a potentially revolutionary tool in clinical practice. The language model has shown promise in assisting physicians in clinical decision-making and formulating personalized therapeutic strategies [17]. However, the existing evidence on ChatGPT's uses in heart failure is limited to a small number of observational studies demonstrating reasonable performance. To date, no randomized controlled trials have been conducted, and there is no documented evidence of its impact on reducing patient admissions or heart failure events. The aim of this systematic review is to critically appraise and synthesize the available evidence on ChatGPT in heart failure. Finally, we highlight the current ethical challenges in adopting ChatGPT technology.

## 2. Methods

### 2.1. Study Design and Reporting Guidelines

This study is a systematic review of original studies and follows the preferred reporting items for systematic reviews and meta-analyses (PRISMA) reporting guidelines. Our systematic review was registered on PROSPERO in August 2025.

### 2.2. Search Strategy

A comprehensive literature search was performed across three major electronic databases: PubMed, Embase, and Cochrane Library from their inception through July 2025, aiming to capture all relevant publications without temporal limitations. The search strategy employed a combination of controlled vocabulary (such as Medical Subject Headings [MeSH] terms where applicable) and free-text keywords to maximize sensitivity. Key search terms were structured as follows: ("ChatGPT" OR "GPT" OR "Generative Pre-trained Transformer") AND ("heart failure" OR "congestive heart failure" OR "HF" OR "CHF"). Boolean operators (AND/OR) were used to refine and broaden the query, with truncation and wildcard symbols applied as needed to account for variations in terminology. No language restrictions were imposed to promote inclusivity and avoid potential bias from excluding non-English studies. Additionally, to identify any overlooked publications, the reference lists of all included studies, as well as pertinent narrative reviews and related articles, were manually hand-searched. Grey literature sources, including conference proceedings and preprint servers, were also screened for completeness. The full search strings for each database are provided in the Supporting Information (Appendix S1).

### 2.3. Inclusion and Exclusion Criteria

The inclusion criteria were as follows:

Published studies demonstrating the current or future use and future potential of ChatGPT in heart failure

ChatGPT's role was considered relevant if it referred to one of the following three domains: education and question-answering, readability enhancement and clinical applications.

Publications relating to both clinical use and academic use are eligible for inclusion

Published in the English language.

The exclusion criteria were as follows:

Abstract-only publications

Studies failing to discuss or denote ChatGPT in heart failure.

#### *2.4. Study Selection, Data Extraction and Critical Appraisal*

A database was created using EndNote X9 (The EndNote Team, Clarivate, Philadelphia, 2013). Abstracts were screened by two independent reviewers (RD and CM) based on inclusion/exclusion criteria, focusing on three domains: ChatGPT's enhancement of patient education and management; factual accuracy of outputs, and comparison to clinical standards. Duplicates were removed, and discrepancies resolved through discussion with a third reviewer (HCT), excluding articles upon agreement. Full texts were evaluated by two reviewers for eligibility, with reference lists hand-searched for additional studies. Data extraction followed the PICOTS framework (Population, Intervention, Comparator, Outcomes, Timing, Setting) using Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia; www.covidence.org). Conflicts were resolved via discussion, with final decisions by the senior author.

A critical appraisal of the methodological quality and risk of bias of the included studies was not performed. There is currently no risk of bias (ROB) tool specific for ChatGPT and as there is no true population ROB cannot be applied to these studies.

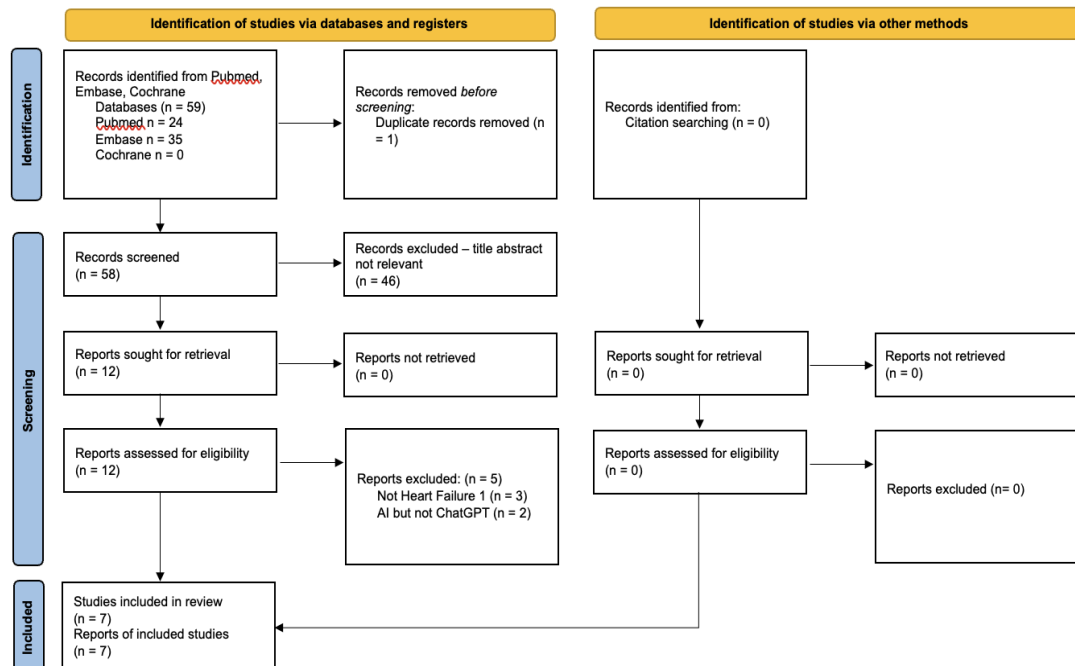
#### *2.5. Synthesis*

A thematic synthesis approach was used to group study findings into three categories: education and question-answering, readability enhancement, and clinical applications. Readability enhancement was defined as the simplification and rephrasing of complex patient education materials to improve accessibility and comprehension while maintaining accuracy. It was measured using tools like the Flesch-Kincaid grade level scores, Patient Education Materials Assessment Tools (PEMAT) readability and actionability percentages, and assessments of word difficulty. Meta-analysis was not conducted due to significant heterogeneity in study designs, interventions, and outcome measures.

### **3. Results**

The PRISMA flow diagram is represented below:

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources



Characteristics of included studies are summarized in Table 1.

**Table 1.** Methodological characteristics of included studies.

Study	Sample Size	Intervention	Comparator	Key Outcomes
Dimitriadis et al. (2024)	47 questions	ChatGPT-4	None (observational)	100% accuracy in key areas such as lifestyle advice, medication mechanisms, and symptom management
King et al. (2024) (appropriateness)	107 questions	GPT-3.5 and GPT-4	GPT-3.5 vs. GPT-4	98.1% appropriateness for GPT-3.5, with occasional hallucinations* (1.9%)
Anaya et al. (2024)	12 questions	ChatGPT-3	Leading US institutes (ACC, AHA, HFSA)	78% actionability score, with competitive 75% PEMAT readability but lowest actionability among compared to materials
Workman et al. (2024)	1999 snippets + 102 synthetic	ChatGPT-4 ZSL	ML and rule-based	95% F1 score for ZSL, outperforming baselines
Kozaily et al. (2024)	30 questions	ChatGPT-3.5 and Bard	ChatGPT vs. Bard	90% appropriateness for ChatGPT
King et al. (2024) (readability)	143 PEMs	GPT-4	Original materials	Improved readability to 6th-7th grade level

Bhupathi et al. (2024)	21 questions (10 HF, 11 PDA)	ChatGPT-3.5	AHA/ACC guidelines	Mean accuracy 5.4/6, 83.75% PEMAT-P understandability
------------------------	------------------------------	-------------	--------------------	---

**Table 2.** Evaluation Methods, Metrics, and Limitations of Included Studies.

Study	ChatGP T Version	Evaluation Method	Metrics Used	Limitations Noted
Dimitriadis et al. (2024)	GPT-4	Expert assessment by cardiologists on accuracy and comprehensiveness	Accuracy (%), comprehensiveness	Less comprehensive; limited to observational design
King et al. (2024) (appropriateness)	GPT-3.5 and GPT-4	Graded by board-certified cardiologists using 4-point scale (comprehensive to incorrect)	Appropriateness (%), reproducibility (%), comprehensive knowledge (%)	Occasional hallucinations* (1.9%); small sample of questions; no patient outcomes
Anaya et al. (2024)	GPT-3	Blind assessment with PEMAT; readability calculators (Flesch-Kincaid, etc.)	PEMAT readability/actionability (%), grade level, word difficulty (%)	Longer responses at higher educational levels; lower actionability; observational only
Workman et al. (2024)	GPT-4	Zero-shot learning with prompt engineering; compared to ML/rule-based baselines	Precision (%), recall (%), F1 score (%)	Reliance on synthetic snippets; no real EHR validation; prompt sensitivity
Kozaily et al. (2024)	GPT-3.5	Expert evaluation by HF specialists; consistency across runs	Appropriateness (%), consistency (%)	Inadequacy in advanced topics; heterogeneity in comparators (Bard); small question set
King et al. (2024) (readability)	GPT-4	Expert grading for accuracy/comprehensiveness; readability scores	Flesch-Kincaid grade level, accuracy (%),	Institutional materials bias; no long-term

			comprehensiveness increase (%)	impact assessment
Bhupathi et al. (2024)	GPT-3.5	Likert scales for accuracy/completeness; PEMAT for understandability	Accuracy score (mean/6), completeness (mean/3), PEMAT understandability (%)	Slightly lower accuracy for rarer conditions; no RCTs; potential data abundance bias

American College of Cardiology (ACC), American Heart Association (AHA), Heart Failure Society of America (HFSA), Patient Education Materials Assessment Tool (PEMAT), Zero-Shot Learning (ZSL), Machine Learning (ML), F1 score (F1), and Generative Pre-trained Transformer (GPT). \* In the context of large language models like ChatGPT, hallucinations refer to instances where the AI generates plausible-sounding but factually incorrect, fabricated, or nonsensical information, often due to limitations in training data or pattern-matching without true understanding.

## 4. Discussion

This systematic review synthesizes evidence from seven observational studies on ChatGPT's applications in heart failure (HF), including applications relating to patient education, question-answering, readability enhancement, and symptom extraction from electronic health records (EHRs). These results align with emerging narrative reviews on ChatGPT in HF [16], which highlight its potential in personalized education and adherence support, but underscore gaps in systematic validation that this review addresses.

### 4.1. Key Findings and Thematic Synthesis

Thematically, the included studies cluster into three domains: patient education and question-answering [29–31,33,35], readability enhancement [31,34], and clinical documentation/symptom extraction [32].

In education and question-answering, ChatGPT provided accurate, empathetic responses that could enhance patient understanding and self-care, with appropriateness rated 90-100% across HF topics like symptoms, lifestyle modifications, and medication management [29,30,33,35]. Dimitriadis et al. (2024) reported 100% accuracy for ChatGPT-4 on 47 common HF questions, excelling in lifestyle advice, medication mechanisms, and symptom recognition. Accuracy was evaluated by two researchers who individually assessed the similarity, relevance, and reliability of responses based on the latest published guidelines for Heart Failure, with overall evaluation by the study's primary supervisor [29]. Kozaily et al. (2024) noted ChatGPT's edge over Bard, Google's AI chatbot, in diagnosis and prognosis but weaknesses in advanced areas like device therapy [33]. Bhupathi et al. (2024) showed better accuracy and completeness for HF than rarer conditions like patent ductus arteriosus (PDA), attributed to greater training data, though using a standard measurement tool, the Patient Education Materials Assessment Tool, material understandability was higher for PDA than for HF information [35]. Anaya et al. (2024) indicated comparable PEMAT readability but lower actionability, when compared with pre-existing patient educational materials from their institution [31]. These comparisons highlight ChatGPT's versatility but underscore limitations in advanced or less common topics.

When asked to enhance readability, ChatGPT simplified complex texts but risked oversimplification. Anaya et al. (2024) reported improvements in actionability through simplification of medical terminology [31], while King et al. (2024) used GPT-4 to rephrase 143 institutional patient education materials (PEMs), reducing Flesch-Kincaid grade levels, maintaining 100% accuracy, and

increasing comprehensiveness in 23% of cases [34]. Anaya et al. [31] also found that ChatGPT-3 responses to HF frequently asked questions were at undergraduate levels, with higher difficult word percentages than ACC materials, yet achieved a PEMAT readability score exceeding the AHA's. However, actionability was lower due to less effective prompts for behavior change [31]. In comparison, King et al. (2024) lowered Flesch-Kincaid scores substantially for institutional HF PEMS but noted risks like loss of technical nuance [34]. This domain contrasts with education and question-answering by focusing on text refinement rather than original content creation.

In symptom extraction, Workman et al. (2024) used ChatGPT-4 to identify heart failure symptoms from simulated electronic health record notes. They applied a "zero-shot" approach, where the AI relies on its general knowledge without special training, and used prompt-engineering for better results. This achieved scores of 90% on precision and 100% on recall, outperforming traditional machine learning methods, which scored 65.5% (F1), without pre-labeled data [32]. Unlike the patient-focused tools for education and readability, this shifts to internal clinical tasks, using ChatGPT's pre-trained knowledge without fine-tuning.

#### 4.2. Implications for Heart Failure Care

ChatGPT addresses HF barriers like low health literacy and poor adherence [11] through conversational simplification [31], thereby building self-efficacy in weight and symptom monitoring, as well as fluid and dietary restriction [16]. Narrative reviews position it as a virtual assistant for diet, exercise, and coping [16,17], which could potentially reduce HF readmissions, which are strongly associated with low health literacy rates [11]. Adherence gains from reminders and explanations align with guideline emphasis on engagement [12]. EHR symptom extraction enables real-time phenotyping and wearable integration for proactive care [16,32]. Unlike the report generation capabilities being employed in radiology (50-100% accuracy [36]), the predominant benefits of ChatGPT in HF care may come in the form of patient-facing self-management tools.

ChatGPT could promote fairness by providing low-cost access to underserved groups [16], but biases such as its English focus might increase inequalities [16,36]. Adapting it for multiple languages is vital, given how social factors affect heart failure outcomes [16].

#### 4.3. Drawbacks and Limitations of ChatGPT in HF Care

While ChatGPT has demonstrated reasonable performance in observational studies, several critical drawbacks must be considered. This is especially relevant for HF patients, whose care can be finely balanced, with even small changes in medications and behaviours potentially leading to decompensations. Current evidence suggests that care led by general physicians, and even general (non-HF specialist) cardiologists, in comparison with care led by specialist HF clinicians, may lead to worse outcomes [37]. Although a 1.9% hallucination rate may appear modest [30], its potential to cause adverse outcomes, such as incorrect medication administration or nonadherence to guideline-directed lifestyle recommendations, could have significant clinical implications for heart failure patients. At a population level, this seemingly small rate could translate into substantial harm, as a 1.9% effect, whether beneficial or harmful, can affect thousands in large patient populations. For context, sacubitril/valsartan (Entresto) demonstrated an absolute risk reduction of 2.8% in all-cause mortality in the PARADIGM-HF trial, a finding that has been widely acclaimed and has markedly influenced contemporary heart failure guidelines and practice [39]. This 2.8% risk reduction seen in PARADIGM-HF is of a magnitude comparable to the 1.9% hallucination rate.

Furthermore, a key strength of ChatGPT lies in its broad accessibility and generally free availability. Much of its current accuracy derives from training on extensive internet datasets, which were largely unrestricted during its development. However, an increasing number of academic journals and professional societies are now implementing protections to prevent unauthorized use of their content for AI training, requiring licenses for such purposes. For instance, the European Society of Cardiology (ESC) has explicitly reserved rights under EU Directive 2019/790, opting out of text and data mining for AI development in their guidelines. This shift could profoundly impact the

application of ChatGPT and other large language models in healthcare if access to the newest evidence-based content is no longer readily available for training. Consequently, patients relying on such models for health information may unknowingly receive compromised, outdated, or non-guideline-directed medical advice.

#### 4.4. Strengths and Limitations of the Evidence

Strengths of the evidence include consistent accuracy across versions and validated readability gains, supported by expert grading. Limitations include observational designs, small samples, and lack of patient outcomes, restricting generalizability and mirroring radiology's gaps [36]. Narrative reviews offer context but lack rigor; this review provides structured quality appraisal. Broader issues involve non-deterministic outputs (44.9% consistency [13]), outdated training risking misinformation [16], ethical concerns (privacy, liability [36]), data cut-offs, biases needing diverse sets, and hallucinations requiring tools like Retrieval-Augmented Generation [16].

#### 4.5. Ethical Considerations

Ethical challenges in deploying ChatGPT for HF include liability, where responsibility for harmful advice could fall on developers, providers, or clinicians amid ambiguous regulations, potentially leading to legal disputes. Financial costs from lawsuits might burden institutions, deterring adoption without evolved insurance models. Contradictions with physician advice could cause confusion, non-adherence, or delays, exacerbating decompensations in HF patients, as seen in case reports of life-threatening misinformation [38]. This may depersonalize care, erode trust, and undermine the doctor-patient relationship. All of these questions are potential barriers to implementation which have not yet been adequately addressed in the Western world and beyond.

Broader ethical concerns in AI for healthcare include data privacy and security, as Large Language Models (LLMs) like ChatGPT process sensitive patient queries that could inadvertently breach General Data Protection Regulation (GDPR) if not properly anonymized or if data is used for model retraining without consent. Bias is another major issue: these models, trained on vast internet datasets, may perpetuate disparities by underperforming for underrepresented groups, such as in ethnic minorities or low-income HF patients, leading to inequitable care outcomes. Transparency poses challenges due to the "black-box" nature of LLMs, making it difficult for clinicians to understand or explain AI-generated advice, which contrasts with evidence-based medicine's emphasis on verifiable reasoning. Informed consent is crucial; patients must be explicitly told they are interacting with an AI, not a human expert, to avoid deception and ensure autonomous decision-making. Finally, equity in access remains a concern; while ChatGPT's low-cost availability is a strength, digital divides, such as a lack of internet or tech literacy in elderly HF populations, could widen health disparities unless mitigated through inclusive design and multilingual support.

#### 4.6. Future Directions

Randomised trials evaluating the effect of ChatGPT versus standard education on patient outcomes including adherence and readmission rates are indicated. Regulatory frameworks, clinician training [36], and multilingual versions are essential for global HF management [17]. Longitudinal studies on engagement and cost-effectiveness, plus refined models for validity, bias, and ethics, will reinforce its utility [16].

## 5. Conclusions

In conclusion, ChatGPT shows considerable promise in improving heart failure management through enhanced patient education, accurate question-answering, improved readability of materials, and efficient symptom extraction. With accuracy rates exceeding 90% in most applications and significant readability gains, ChatGPT addresses critical barriers like low health literacy and adherence, potentially reducing HF's global burden. However, risks of misinformation and ethical

concerns necessitate cautious integration. Future research should prioritize randomized trials, real-world validations, and bias mitigation to harness ChatGPT's full potential, ensuring equitable, safe, and effective AI-driven HF care.

**Author Contributions:** Conceptualization, R.S.D, H.C.T., C.M. ; methodology, R.S.D, J.H., H.C.T., C.M., R.W., J.W. ; software, R.S.D., H.C.T.; validation, R.D., H.C.T., C.M. ; formal analysis, R.D., H.C.T., C.M., J.H., R.W., J.W. ; investigation, R.D., H.C.T., C.M., J.H., R.W., J.W. ; resources, R.D., H.C.T., C.M., J.H., R.W., J.W., J.McC. ; data curation, R.D., H.C.T., C.M., J.H., R.W., J.W. ; writing—original draft preparation, R.S.D, J.H., J.McC., H.C.T., C.M., R.W., J.W. ; writing—review and editing, R.S.D, J.H., J.McC., H.C.T., C.M., R.W., J.W., C.M., K.M., K.McD.; visualization, R.S.D, J.H., J.McC., H.C.T., C.M., R.W., J.W., C.M., K.M., K.McD. ; supervision, K.McD. ; project administration, R.S.D, H.C.T., J.H. ; funding acquisition, not applicable. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data analyzed in this systematic review are derived from publicly available studies cited in the references section. No new primary data were generated; all extracted data are contained within the article and its supplementary materials. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the United States. *Circ Heart Fail.* 2013;6(3):606-619. doi:10.1161/CIRCHEARTFAILURE.113.000119
2. Savarese G, Becher PM, Lund LH, et al. Global burden of heart failure: a comprehensive and updated review of epidemiology. *Cardiovasc Res.* 2023;119(6):1452-1469. doi:10.1093/cvr/cvac013
3. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the Impact of Heart Failure in the United States: A Policy Statement From the American Heart Association. *Circ Heart Fail.* 2013;6:606-619.
4. Dunlay SM, Shah ND, Shi Q, et al. Lifetime costs of medical care after heart failure diagnosis. *Circ Cardiovasc Qual Outcomes.* 2011;4(1):68-75. doi:10.1161/CIRCOUTCOMES.110.957225
5. Khan MS, Sreenivasan J, Lateef N, et al. Trends in 30- and 90-day readmission rates for heart failure. *Circ Heart Fail.* 2021;14(5):e008335. doi:10.1161/CIRCHEARTFAILURE.121.008335
6. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure. *Circulation.* 2022;145(18):e895-e1032. doi:10.1161/CIR.0000000000001063
7. McCullough PA, Mehta HS, Barker CM, et al. Mortality and guideline-directed medical therapy in real-world heart failure patients with reduced ejection fraction. *Clin Cardiol.* 2021;44(9):1192-1198. doi:10.1002/clc.23664
8. Ross JS, Chen J, Lin Z, et al. Recent national trends in readmission rates after heart failure hospitalization. *Circ Heart Fail.* 2010;3(1):97-103. doi:10.1161/CIRCHEARTFAILURE.109.885210
9. Gautam N, Ghanta SN, Clausen A, et al. Contemporary applications of machine learning for device therapy in heart failure. *JACC Heart Fail.* 2022;10(9):603-622. doi:10.1016/j.jchf.2022.06.002
10. Khan MS, Arshad MS, Greene SJ, et al. Artificial intelligence and heart failure: a state-of-the-art review. *Eur J Heart Fail.* 2023;25(9):1507-1525. doi:10.1002/ehf.2988
11. Fabbri M, Yost KJ, Finney Rutten LJ, et al. Health literacy and outcomes in patients with heart failure: a prospective community study. *Mayo Clin Proc.* 2018;93(1):9-15. doi:10.1016/j.mayocp.2017.09.018
12. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42(36):3599-3726. doi:10.1093/eurheartj/ehab368
13. Funk PF, Hoch CC, Knoedler S, et al. ChatGPT's response consistency: a study on repeated queries of medical examination questions. *Eur J Investig Health Psychol Educ.* 2024;14(3):657-668. doi:10.3390/ejihpe14030043

14. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097
15. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* 2016;355:i4919. doi:10.1136/bmj.i4919
16. Ghanta SN, Dahal R, Vellanki S, et al. Applications of ChatGPT in heart failure prevention, diagnosis, management, and research: a narrative review. *Diagnostics (Basel).* 2024;14(21):2372. doi:10.3390/diagnostics14212372
17. Sharma A, Gupta R, Patel N, et al. Exploring the role of ChatGPT in cardiology: a systematic review of the current literature. *Cureus.* 2024;16(5):e59512. doi:10.7759/cureus.59512
18. Sarraju A, Bruemmer D, van Iterson E, et al. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *JAMA.* 2023;329(10):842-844. doi:10.1001/jama.2023.1044
19. 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). *Acta Cardiol.* 2024;79(3):358-366. doi:10.1080/00015385.2024.2321555
20. Hirokawa T, Kawamura R, Harada Y, et al. ChatGPT-generated differential diagnosis lists for complex case-derived clinical vignettes: diagnostic accuracy evaluation. *JMIR Med Inform.* 2023;11:e48808. doi:10.2196/48808
21. Roosan D, Padua P, Khan R, et al. Effectiveness of ChatGPT in clinical pharmacy and the role of artificial intelligence in medication therapy management. *J Am Pharm Assoc (2003).* 2024;64(2):422-428. doi:10.1016/j.japh.2023.11.023
22. Alanzi TM. Impact of ChatGPT on teleconsultants in healthcare: perceptions of healthcare experts in Saudi Arabia. *J Multidiscip Healthc.* 2023;16:2309-2321. doi:10.2147/JMDH.S421244
23. Rawashdeh B, Kim J, AlRyalat SA, et al. ChatGPT and artificial intelligence in transplantation research: is it always correct? *Cureus.* 2023;15(7):e42150. doi:10.7759/cureus.42150
24. Elyoseph Z, Hadar-Shoval D, Asraf K, Lvovsky M. ChatGPT outperforms humans in emotional awareness evaluations. *Front Psychol.* 2023;14:1199058. doi:10.3389/fpsyg.2023.1199058
25. Guevara M, Chen S, Thomas S, et al. Large language models to identify social determinants of health in electronic health records. *NPJ Digit Med.* 2024;7(1):6. doi:10.1038/s41746-023-00970-0
26. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst.* 2023;47(1):33. doi:10.1007/s10916-023-01925-4
27. Nakaya Y, Higaki A, Yamaguchi O. ChatGPT's ability to classify virtual reality studies in cardiology. *Eur Heart J Digit Health.* 2023;4(2):141-142. doi:10.1093/ehjdh/ztad009
28. Fijačko N, Prosen G, Abella BS, et al. Can novel multimodal chatbots such as Bing Chat Enterprise, ChatGPT-4 Pro, and Google Bard correctly interpret electrocardiogram images? *Resuscitation.* 2023;193:110009. doi:10.1016/j.resuscitation.2023.110009
29. Dimitriadis F, Alkagiet S, Tsigkriki L, et al. ChatGPT and patients with heart failure. *Angiology.* 2024. doi:10.1177/00033197241238403
30. King RC, Samkari JS, Haquang J, et al. Appropriateness of ChatGPT in answering heart failure related questions. *Heart Lung Circ.* 2024;33(9):1314-1318. doi:10.1016/j.hlc.2024.03.005
31. Anaya F, Prasad R, Bashour M, et al. Evaluating ChatGPT platform in delivering heart failure educational material: a comparison with the leading national cardiology institutes. *Curr Probl Cardiol.* 2024;49(10):102797. doi:10.1016/j.cpcardiol.2024.102797
32. Warkins TE, Leese J, Ouyang D, et al. ChatGPT-4 zero-shot learning extraction of heart failure signs and symptoms from electronic health records. *Prog Cardiovasc Dis.* 2024;87:47-49. doi:10.1016/j.pcad.2024.09.001
33. Kozaily E, Geagea M, Akdogan ER, et al. Accuracy and consistency of online large language model-based artificial intelligence chat platforms in answering patients' questions about heart failure. *Int J Cardiol.* 2024;408:132115. doi:10.1016/j.ijcard.2024.132115

34. King RC, Srinivasan N, Peng Y, et al. Improving the readability of institutional heart failure-related patient education materials using GPT-4: observational study. *JMIR Cardio*. 2024;8:e48817. doi:10.2196/48817
35. Bhupathi M, Mediboina A, Mehweish I, et al. Assessing information provided by ChatGPT: heart failure versus patent ductus arteriosus. *Cureus*. 2024;16(6):e60365. doi:10.7759/cureus.60365
36. Temperley HC, O'Sullivan NJ, Mac Curtain BM, et al. Current applications and future potential of ChatGPT in radiology: a systematic review. *J Med Imaging Radiat Oncol*. 2024;68(3):250-260. doi:10.1111/1754-9485.13621
37. Warkentin J, Ezekowitz JA, White J, et al. Care gaps in heart failure management: impact of subspecialty care on outcomes. *J Card Fail*. 2024; doi:10.1016/j.jchf.2024.06.013
38. Saenger JA, Hunger J, Boss A, et al. Delayed diagnosis of a transient ischemic attack caused by ChatGPT. *Wien Klin Wochenschr*. 2024;136(7-8):236-238. doi:10.1007/s00508-024-02329-1
39. McMurray JJV, Packer M, Desai AS, et al. Angiotensin–Neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014;371(11):993-1004. doi:10.1056/NEJMoa1409077.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.