ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

REVIEW ARTICLE

The role of large language models in induced pluripotent stem cell-derived cardiomyocytes research and clinical translation

Dhienda C. Shahannaz[1,2], Tadahisa Sugiura[2*], and Brandon E. Ferrell[2]

[1]Department of Medicine, Medical Education and Research Institute, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia

[2]Department of Cardiothoracic and Vascular Surgery, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York, United States of America

(This article belongs to the *Special Issue: Exploring the Potential of Large Language Models (ChatGPT) in Cardiovascular Disease Management*)

Abstract

Background: Induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs) are redefining cardiovascular regenerative medicine, yet challenges in differentiation fidelity, functional maturation, and scalable production restrain their full clinical potential. **Aim:** This review evaluates the pioneering integration of large language models (LLMs)—including GPT-4, BioGPT, and BioMedLM—into iPSC-CM research and translational therapeutics, with a focus on advancing precision, efficiency, and patient-specific care. **Methods:** Structured searches across biomedical and artificial intelligence-focused databases were conducted to map how LLMs augment literature mining, experimental design, multi-omics integration, and clinical translation, including personalized therapy prediction and drug safety assessment. **Results:** LLMs demonstrably surpass traditional tools in identifying gene-phenotype links, refining clustered regularly interspaced short palindromic repeats-based differentiation protocols, and merging patient-level datasets with iPSC-CM outputs. Limitations include model interpretability, reproducibility across genetically diverse populations, and ethical considerations regarding data privacy and bias. **Conclusion:** Despite these barriers, early translational applications demonstrate that LLMs can accelerate hypothesis generation, optimize laboratory-to-clinic pipelines, and enable high-fidelity, patient-specific cardiomyocyte modeling. **Relevance for patients:** The synergy of LLM intelligence and iPSC-CM biology has the potential to deliver safer, more effective, and deeply personalized regenerative cardiac therapies—moving the field closer to truly bespoke heart repair.

Keywords: Artificial intelligence; Large language models; Biomedical natural language programs; Induced pluripotent stem cells; Cardiac regenerative medicine

***Corresponding author:**
Tadahisa Sugiura
(tsugiura@montefiore.org)

Citation: Shahannaz DC, Sugiura T, Ferrell BE. The role of large language models in induced pluripotent stem cell-derived cardiomyocytes research and clinical translation. *J Clin Transl Res*. 2025;11(5):4-28.
doi: 10.36922/JCTR025230026

Received: June 3, 2025

Revised: July 20, 2025

Accepted: August 7, 2025

1. Introduction

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide,[1] necessitating continuous advancements in therapeutic strategies. According to the World Health Organization, CVDs account for 32% of global deaths,[2] with Indonesia recording 651,481 CVD-related deaths (38.2%),[3,4] the United States 957,455 deaths (35.7%),[5] and Japan 372,483 deaths (28.0%).[6] In high-performing healthcare systems, CVD-related fatalities can be reduced below 20% through advanced preventive strategies, early intervention, and innovative therapeutic solutions.[7-10] One such breakthrough is the development of induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs),[11] which offer potential applications in disease modeling, drug testing, and cardiac tissue engineering. However, challenges persist, including variability in differentiation protocols, limited functional maturation, and obstacles to large-scale clinical application.[12]

The rise of artificial intelligence (AI)[13] and large language models (LLMs)[14] has opened new avenues to accelerate iPSC-CM research and clinical translation. Historically, AI-driven innovations have reshaped cardiovascular medicine.[15] Machine learning has enhanced diagnostic imaging,[16] refined risk prediction models,[17] and optimized surgical planning in cardiothoracic procedures.[18] More recently, LLMs, such as ChatGPT (OpenAI), DeepSeek (Hangzhou DeepSeek AI Company), Bard (Google AI), and GROK (xAI),[19] have revolutionized biomedical research by enabling large-scale data analysis,[20,21] optimizing differentiation strategies,[22-25] and predicting patient-specific responses to regenerative therapies.[26,27] Since its release on November 30, 2022, ChatGPT reached one million users in just five days—far surpassing the growth of platforms, such as Facebook, which took nearly 10 months to reach the same milestone. Remarkably, ChatGPT has also demonstrated performance comparable to a 3rd-year medical student on the National Board of Medical Examiners assessments and passed the United States Medical Licensing Examination Step exams,[25] underscoring its potential to contribute meaningfully to high-accuracy domains, such as stem cell-based cardiovascular research.[28] Despite these advancements, current applications of LLMs in iPSC-CM research remain underexplored, with key gaps in long-term validation, reproducibility, and standardization.

This review critically examines the evolving role of LLMs in iPSC-CM research and translation. Through targeted analysis of current literature, it explores how LLM-based frameworks can enhance differentiation strategies, uncover functional biomarkers, and bridge lab-based insights with clinical application, laying a foundation for more scalable and precise cardiovascular regenerative solutions.

2. Methods

To synthesize a comprehensive view of LLM applications in iPSC-CM research and clinical translation, a narrative review methodology was adopted. Relevant studies were identified through targeted searches across PubMed, Google Scholar, arXiv, and Web of Science using combinations of the following terms: "LLM," "large language model," "iPSC-CM," "induced pluripotent stem cell," "cardiomyocyte differentiation," "regenerative cardiology," "cardiotoxicity," "CRISPR screen," "single-cell RNA-seq," and "deep learning." Additional queries incorporated more specific phrases, including "LLM in clinical genomics," "cardiac lineage specification," "iPSC-CM drug screening," "electronic health records," "BioBERT," "BioMedLM," and "protein structure prediction."

Inclusion criteria comprised: (i) peer-reviewed articles, preprints, or white papers describing the use of AI or LLMs in cardiovascular, stem cell, or regenerative research; (ii) studies involving iPSC-CMs in disease modeling, drug screening, or translational applications; and (iii) sources published in English from 2018 onward to reflect the advent of transformer-based architectures.

Exclusion criteria included: (i) studies not involving cardiovascular applications or not using iPSC-CMs; (ii) non-AI-based reviews or purely theoretical discussions without applied methodology; and (iii) articles lacking relevance to clinical translation or omics-driven discovery.

No strict limitations on publication types were imposed, allowing the inclusion of preclinical, computational, and translational studies. Approximately 150 sources were screened, with 45 core references included in the final synthesis based on thematic relevance, methodological quality, and impact on the evolving role of LLMs in cardiovascular regenerative medicine.

3. Results and discussion

Although structured as a narrative review, we integrate comparative insights and propose a scaffolding for future benchmarking protocols in iPSC-CM applications of LLMs.

3.1. Summary of key findings

LLMs, such as ChatGPT, are increasingly integrated into clinical and research workflows, supporting peer discussions, complex decision-making, and interdisciplinary planning. In cardiothoracic contexts, they assist with surgical preparation, data analysis, literature synthesis, and knowledge translation—supporting expertise sharing, collaborative planning, and innovation in iPSC-CM research.[24,26] The visual overviews of this pipeline are shown in Figures 1 and 2.

Figure 1 illustrates the progressive specialization of AI tools–from general-purpose AI to clinical personalization through LLM-enhanced multi-omics modeling customized for cardiac regenerative contexts. Figure 2 outlines the

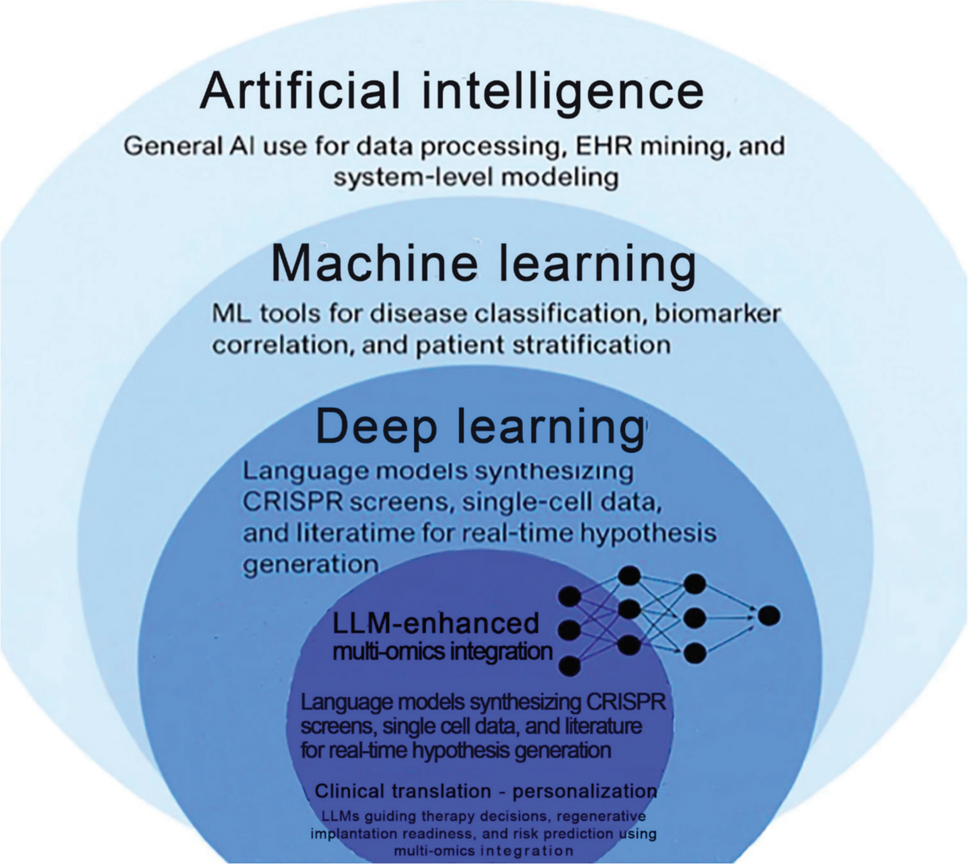

Figure 1. Layered AI-LLM integration in iPSC-CM research and clinical translation. Image created by the authors.
Abbreviations: AI: Artificial intelligence; CRISPR: Clustered regularly interspaced short palindromic repeats; EHR: Electronic health record; LLM: Large language model; ML: Machine learning.

integration of LLMs across the five-phase iPSC-CM research and clinical translation workflow: (i) literature mining and knowledge extraction: LLMs, such as BioGPT and ChatGPT, summarize protocols, annotate biomarkers, and extract disease-gene associations from biomedical corpora, (ii) target and pathway discovery: deep generative models, such as BioMedLM and AlphaMissense, prioritize variants, and signaling axes (e.g., *PGC1α* and *SIRT3*) relevant to mitochondrial maturation, (iii) *In silico* modeling of molecular interactions: Structure predictors (AlphaFold and RoseTTAFold) map mutation-driven conformational changes, while JAX and PyTorch simulate cardiomyocyte differentiation trajectories, (iv) functional testing in iPSC platforms: AI-guided experiment planners optimize clustered regularly interspaced short palindromic repeat (CRISPR) screens and electrophysiological readouts using tools like scGPT and DeepChem, and (v) clinical translation and risk prediction: multimodal fusion of omics + electronic health records (EHR) data supports transplant safety scoring, arrhythmia prediction, and therapy personalization through platforms, such as REALM and CardioGenAI. Arrows denote LLM-facilitated knowledge flow. Annotations highlight model-specific tasks. This framework emphasizes interpretability, reproducibility, and predictive fidelity across patient-specific and population-scale applications. This figure also illustrates the methodological diversity across international studies, enabling comparison between molecular-targeting and clinical-triage LLM use cases.

LLMs have become key tools in modeling iPSC-CM maturation. They also support clinical translation by handling complex, multi-layered datasets. At the molecular level, key maturation hallmarks, such as sarcomere alignment, T-tubule formation, and mitochondrial biogenesis, are increasingly understood through integration of single-cell transcriptomics,[29,30] epigenomic atlases,[31-35] and proteomic datasets. In particular, mitochondrial maturation has gained central focus, as iPSC-CMs transition from a glycolytic, fetal-like metabolic profile to one reliant on mitochondrial oxidative phosphorylation, characteristic of mature cardiomyocytes. Recent studies, including a study in Spain by Zamora-

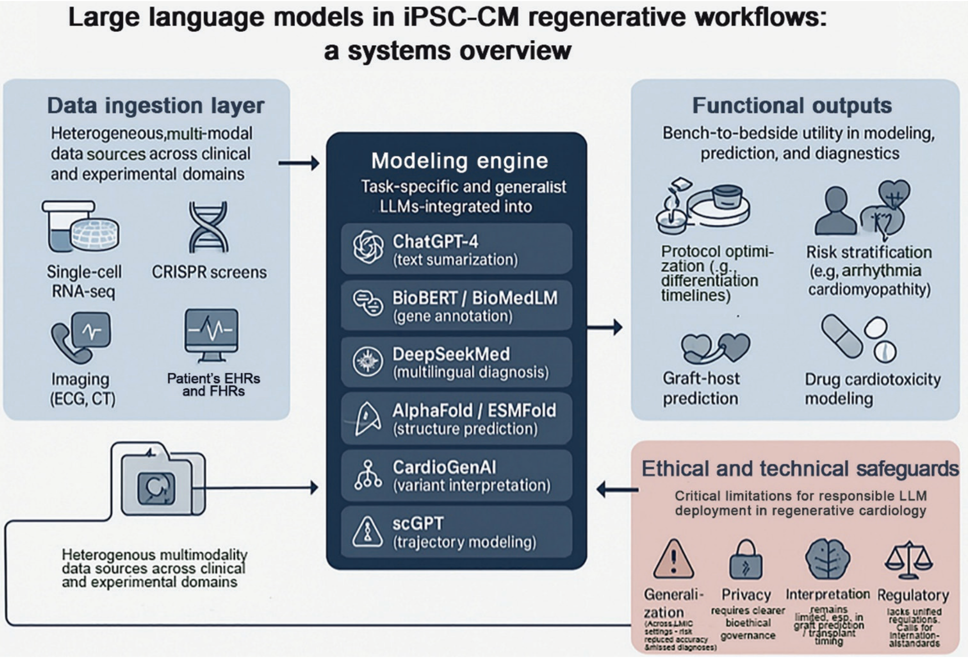

Figure 2. LLM-Augmented iPSC-CM Research and Translation Pipeline. Arrows denote the flow of data and knowledge; annotations highlight the model's functions and limitations. Image created by the authors.
Abbreviations: CRISPR: Clustered regularly interspaced short palindromic repeats; CT: Computed tomography; ECG: Electrocardiogram; EHRs: Electronic health records; FHRs: Functional heart readouts; LLM: Large language model; LMIC: Low- and middle-income countries; ML: Machine learning.

Dorta *et al.*,[36] utilized time-resolved metabolomics and CRISPR libraries to trace metabolic reprogramming (e.g., identifying *RTN4IP1* and *ECHS1*), providing a functional contrast to Liu *et al.*'s[37] broader transcriptomic atlas in iPSC-CM maturation. Together, these findings pave the way for using similar CRISPR-based tools to determine whether temporal activation and modulation of *PGC-1α*, *MFN2*, and *SIRT3* can enhance post-transplant integration and functional maturation.

Table 1 summarizes key studies that directly incorporate AI and LLMs into cardiovascular research, highlighting their methodologies, systems used, key findings, and limitations. While this paper follows a narrative review structure, the comparative table serves to clarify specific contributions and gaps across the current literature. Rather than providing a quantitative meta-analysis, it distills representative examples to scaffold the discussion that follows. These studies demonstrate a growing yet uneven integration of LLMs in clinical and experimental cardiology, often limited by a lack of benchmarking, small sample sizes, or conceptual framing without implementation. This underscores the need for more rigorous computational evaluation and real-world application trials.

On the computational front, advancements in LLM programming—including transformer-based architectures and integration with programming libraries, such as PyTorch,[40] JAX,[41,42] and HuggingFace transformers[43]—have enabled more efficient modeling of high-dimensional omics data. LLMs trained on scientific literature, laboratory records, and genomic annotations support the generation of hypotheses, design of protocols, and annotation of maturation-specific expression networks. For example, Google DeepMind's use of reinforcement learning on amino acid-specific datasets, AlphaMissense,[44] combined with LLM-assisted literature mining and streamlining CRISPR-based editing and functional assays, has reconstructed cardiac gene regulatory networks involving *NKX2-5*, *GATA6*, and *MYL2*, providing a systems-level view and enabling efficient mapping of variants affecting these regulators in cardiomyocyte differentiation. Cross-institutional efforts in Japan are now implementing LLM-assisted pipelines in pursuit of their first model suite, which will further impact the iPSC-CM studies and clinical translation while positioning themselves at the forefront of global AI-powered biomedical discovery.[45]

Table 2 compares the traditional workflows in iPSC-CM research and clinical translation using LLM-enhanced workflows. This table compares the evolution of iPSC-CM research with the support of AI. In the traditional workflow, each stage—such as reading papers, designing experiments, and analyzing data—relies heavily on manual labor and human memory. With AI integration, especially language models, tasks become faster, smarter,

Table 1. Comparative evaluation of AI and LLM studies in cardiovascular and biomedical research

Study	Methodological approach	Model system	Key finding	Notable limitation
1. Liu *et al.*[37]	CNN on echocardiograms	Human echo datasets	Outperformed cardiologists in detecting HCM	Reduced accuracy on underrepresented ethnicities
2. Panahiazar *et al.*[38]	Random forest and EHR	Retrospective EHR	Predicted heart failure six months in advance	No external validation across systems
3. Olawade *et al.*[15]	Narrative review of AI in cardiology in general	AI in general	Reviewed current AI trends in cardiology, highlighting LLMs' potential in diagnostics and clinical support	Lacked specific model validation or benchmarking
4. Tolu-Akinnawo *et al.*[16]	Systematic literature review	AI in non-invasive cardiac imaging	Improved cardiac image analysis accuracy, including LLM support in annotation and automation.	Applied heterogeneous validation metrics across studies.
5. Kasartzian and Tsiampalis[17]	Review	ML/AI for cardiac risk prediction	Outperformed traditional risk calculators in CVD risk assessment.	Limited real-world implementation
6. Leivaditis *et al.*[18]	Review	AI in cardiac surgery	LLMs supported surgical planning and patient stratification	Need for data standardization and a regulatory framework
7. Salihu *et al.*[24]	Pilot study	ChatGPT for heart team decision making	ChatGPT enhanced team communication in evaluating severe arctic stenosis cases	Limited by a small sample size and a qualitative nature
8. Ahmed *et al.*[25]	Opinion peace	ChatGPT in cardiothoracic surgery	Outlined the potential of ChatGPT to improve pre-/post-operation patient communication	No empirical data or case application
9. Clark[27]	Perspective	ChatGPT in cardiac surgery and transplantation	Proposed integration of LLMs in education and procedural support	Remained conceptual, no implementation data.
10. Chen *et al.*[21]	Model development	Multi-role ChatGPT framework	ChatGPT can assist in clinical summarization and medical data structuring	Dependent on prompt engineering
11. Iqbal *et al.*[20]	Umbrella review	LLM in healthcare (focus on ChatGPT)	Strong potential for clinical communication, patient interaction, and record summarization	Unresolved hallucination and bias mitigation
12. Pan *et al.*[39]	Computational-expert hybrid pipeline	LLM human integration in EHR	LLMs boost disease detection accuracy when augmented with clinical oversight	Required clinical validation to avoid misclassification

Abbreviations: AI: Artificial intelligence; CNN: Convolutional neural network; CVD: Cardiovascular disease; EHRs: Electronic health records; HCM: Hypertrophic cardiomyopathy; LLMs: Large language models; ML: Machine learning.

and more personalized, from literature scans to clinical predictions. The AI-enhanced approach offers deeper insights, minimizes human bias, and brings research closer to real-world applications with unmatched precision. Clinically, these integrated tools support the refinement of maturation protocols and enhance patient-specific therapeutic planning. By mapping mitochondrial density and electrophysiological maturity across iPSC-CM cohorts, LLMs can identify underdeveloped grafts that are unsuitable for transplantation, ensuring safety and efficacy. This integration of cellular bioenergetics with computational modeling supports strategies for biologically aligned and data-driven cardiac regeneration. This integration provides a framework for advancing translational insight and supporting clinical standardization of regenerative therapies.

Comparative studies have begun to show the superiority of transformer-based models over traditional logistic regression or rule-based natural language processing (NLP) in tasks, such as myocardial infarction identification from clinical notes or gene-phenotype linkage in iPSC-derived platforms. For instance, BioGPT outperformed MetaMap and cTAKES in semantic accuracy when classifying drug-induced arrhythmia mechanisms from biomedical abstracts. The comparative evaluation of LLMs in cardiovascular applications is discussed in Section 3.8.

3.2. Predictive modeling and diagnostics for iPSC-CMs disease modeling and early intervention

LLMs are transforming the landscape of predictive and diagnostic cardiology by integrating patient-level, biomolecular, and physiological datasets (Table 2). In the context of iPSC-CMs, LLMs can forecast disease phenotypes by analyzing genomic instability, ion channel transcriptomics, and electrophysiological aberrancies associated with arrhythmogenic and dilated

Table 2. Traditional versus LLM-enhanced workflows in iPSC-CM research and translation

Research stage	Traditional workflow	LLM-enhanced workflow
1. Literature review	Manual curation across databases and time-consuming filtering of relevant studies, limiting systematic analysis	AI-driven comprehensive reviews of extensive biomedical databases; automated retrieval, summarization, and contextual comparison of numerous papers through NLP
2. Experimental design	Institution-based hypothesis formation, heavily reliant on prior lab protocols and trial-and-error	LLM-assisted generation of precise and testable research questions, AI-assisted hypothesis generation, and protocol optimization based on similar published data
3. Data analysis	Statistical tools-based analysis (e.g., R and SPSS) of omics/electrophysiology, requiring multi-tool integration and specialist knowledge	Integration of multi-omics, phenotypic, and high-resolution imaging datasets. Multimodal integration of scRNA-Seq, CRISPR, proteomics, and imaging through unified AI models
4. Interpretation	Results interpretation by human researchers and subject experts, with potential risk of bias or oversight	Model-driven mechanistic insights with reduced bias; LLMs highlight underreported pathways, variant impact predictions, and potential modifiers
5. Clinical translation	Limited predictive capabilities on patient-specific responses, manual correlation between lab results and clinical outcomes	Personalized modeling of disease states and therapeutics; predictive modeling of therapy outcomes using EHRs, patient genomics, and LLM-generated risk profiles

Abbreviations: AI: Artificial intelligence; CRISPR: Clustered regularly interspaced short palindromic repeats; EHRs: Electronic health records; iPSC-CM: Induced pluripotent stem cell-derived cardiomyocytes; LLMs: Large language models; NLP: Natural language processing; scRNA: Single-cell RNA; SPSS: Statistical Package for the Social Sciences.

cardiomyopathies.[44-46] These models have demonstrated efficacy in predicting calcium-handling dysfunctions, sarcomeric gene disruptions, and metabolic shifts during cardiomyocyte maturation.[47] Clinically, LLM-driven platforms support early identification of myocardial ischemia and hypertrophy by integrating wearable telemetry, EHR-derived hemodynamics,[48] and laboratory markers, such as N-terminal pro-B-type natriuretic peptide,[49,50] troponins, and C-reactive protein.[51] Their ability to continuously learn from cross-institutional datasets allows them to fine-tune treatment decisions—suggesting beta-blocker versus angiotensin-converting enzyme inhibitor therapy in hypertensive heart disease,[52] or even proposing individualized antiarrhythmic strategies based on ion channel mutations (e.g., CardioGenAI).[53,54]

3.2.1. Major AI and LLM tools and platforms

Various AI and LLM platforms are currently integrated into iPSC-CM research pipelines and clinical translation workflows. These tools span from structure prediction and language modeling to real-time diagnostics and simulation:

(i) AlphaFold: A deep learning (DL) system developed by DeepMind that predicts the three-dimensional (3D) structure of proteins. It uses DL algorithms and protein structure databases to accurately determine the folding patterns and spatial arrangements of amino acids in protein sequences. This has revolutionized discovery workflows in structural and molecular biology. Alphafold's exceptional performance in the Critical Assessment of Structure Prediction competition has garnered widespread recognition.[55,56]

(ii) AlphaMissense: AlphaFold-based DL used to model pathogenicity of missense mutations; integrated into LLM pipelines for variant interpretation in cardiac genes, such as *MYL2* and *NKX2-5*[44]

(iii) BioBERT: Biomedical-focused LLM that supports annotation, relation extraction, and hypothesis generation in iPSC-CM molecular modeling and literature mining[57,58]

(iv) BioGPT: Biomedical generative transformer LLM for summarizing research findings, generating hypotheses, and automating insight extraction from omics data[59]

(v) BioMedLM: Biomedical-focused LLM trained on biomedical literature; useful in LLMs tasked with summarizing cardiac differentiation protocols or interpreting biomarker literature[60]

(vi) Cardiogen AI: Developed by BGI Genomics, it is an automated interpretation AI system that links genetic variants to clinical phenotypes in monogenic CVDs. It assists clinicians in diagnosing conditions, such as cardiomyopathies and hypertension, by providing a comprehensive genotype-phenotype database, enhancing precision medicine approaches in cardiology[53,54]

(vii) ChatGPT: General LLM used in literature synthesis, research planning, protocol brainstorming, and peer discussions; not domain-specific but widely integrated in clinical planning[21-27]

(viii) Chemputer: AI-driven chemistry automation aims to revolutionize the field of chemistry by automating and digitizing the chemical synthesis process. The Chemputer system combines robotics, AI, and machine learning to enable the automated design and synthesis of complex molecules. It allows chemists to program and control the synthesis of specific compounds, improving efficiency. The ultimate goal of Chemputer is to accelerate the discovery and development of new compounds for applications in drug discovery, materials science, and beyond[61]

(ix) ClinVar: National Institute of Health (NIH)-owned variant, a public database that archives reports of the relationships between human genetic variants and their clinical significance. It collects submissions from research, labs, clinics, and researchers, helping to classify whether specific variants are benign, pathogenic, or of uncertain significance[62]

(x) DeepChem: DL framework for *in silico* drug modeling and structure-activity prediction; integrated into iPSC-CM cardiotoxicity screening pipelines using LLM-generated compound profiling[63,64]

(xi) DeepSeek-Med: Chinese biomedical LLM initiative; mentioned as a potential future collaborator in international AI consortia for regenerative platforms and clinical translation[19,58]

(xii) Ensembl Genome Browser: A genomic data hub platform that provides integrated, annotated reference genomes for a wide range of species, including humans. It enables researchers to explore genes, variants, regulatory regions, and comparative genomic data. In cardiac research and clinical translation, Ensembl plays a crucial role in identifying genetic mutations and regulatory elements linked to heart diseases[65,66]

(xiii) ESMFold: DL model developed by Meta AI that predicts protein 3D structures directly from amino acid sequences-similar to AlphaFold, but optimized for speed and scalability. It uses LLM principles trained on millions of protein sequences to understand protein folding patterns without relying on multiple sequence alignments. In cardiac research and clinical translation, ESMFold can help predict how mutations in cardiac-related proteins, such as ion channels or sarcomeric proteins, alter their structure and function. This is crucial for understanding diseases, such as cardiomyopathies or channelopathies[69]

(xiv) GEO: NIH-owned high-throughput gene expression and sequencing database repository, such as RNA-seq and microarray results. Researchers submit datasets from various tissues, cell types, and experimental conditions, including cardiac cells and disease models. GEO enables scientists to explore gene regulations in heart disease, development, or drug response. It is used for discovering biomarkers, understanding disease mechanisms, and validating experimental findings in iPSC-derived cardiomyocytes[69]

(xv) GROK: LLM developed by xAI; referenced in the context of AI landscape expansion but not yet applied in iPSC-CM pipelines[19]

(xvi) HuggingFace Transformers: LLM library hub used for implementing transformer-based models, including BioBERT and REALM; provides the backbone for LLM fine-tuning in multi-modal omics data pipelines[43]

(xvii) JAX: High-performance ML framework used in cardiac LLMs for omics modeling and optimization tasks, including protocol efficiency simulations in iPSC-CM studies[41,42]

(xviii) PyTorch: DL library used to build and train LLMs for cardiac modeling, including time-series prediction and transformer network construction[40,70,71]

(xix) REALM: Retrieval-augmented language model combining LLMs with document retrieval—used in EHR mining and real-time diagnostic applications, including arrhythmia detection[72]

(xx) RoseTTAfold (Baker Lab): A DL tool that predicts protein structures from amino acid sequences with high accuracy. It utilizes a three-track neural network to integrate sequence, distance, and coordinate data, enabling rapid modeling of protein structures. In cardiac research, RoseTTAfol aids in understanding the structural implications of genetic mutations associated with heart diseases. By predicting how specific mutations affect protein folding and function, researchers can identify potential targets for therapeutic intervention. This is particularly valuable when experimental structures are unavailable, allowing for exploration of disease mechanisms at the molecular level[56,73-75]

(xxi) scFoundation: LLM foundation model for single-cell data integration; supports high-resolution subtype prediction and cardiac developmental mapping in iPSC-CM pipelines[47]

(xxii) scGPT: Generative LLM tailored for single-cell omics; used in predicting cell fate trajectories, cardiac subtype classification, and transcriptomic modeling[46]

(xxiii) TensorFlow: DL library used for implementing deep learning models, including convolutional neural networks and recurrent neural networks, for cardiac imaging, time-series EHR data, or iPSC-CM signal traces.[70,71,76]

3.2.2. Comparative utility of biomedical LLMs

While Table 2 outlines technical specifications and training corpora across a diverse range of LLMs—from general-purpose models, such as ChatGPT and DeepSeek, to domain-specific engines, such as BioGPT and ClinicalCamel—it is important to highlight their comparative utility in real-world cardiovascular contexts. For instance, BioGPT and PubMedGPT have demonstrated superior term-precision in omics literature mining, especially in identifying gene-regulatory networks relevant to sarcomeric function and cardiac reprogramming. In contrast, DeepSeekMed and DoctorGLM, optimized for multilingual corpora, have outperformed baseline models in extracting phenotypic annotations from iPSC-CM differentiation protocols in both Chinese and English datasets. Experimental benchmarks from Japanese and U.S. institutions have also reported LLM-enhanced accuracy in predicting arrhythmogenic gene clusters and drug-drug cardiotoxicity interactions when integrated with CRISPR screen outputs. These comparative findings support the translational validity of such models, moving them beyond theoretical constructs into tools with tangible experimental and clinical consequences.

LLMs are redefining the diagnostic and predictive capabilities of iPSC-CM platforms by merging computational insight with molecular fidelity. Conventionally, disease modeling using iPSC-CMs has faced challenges in achieving sufficient phenotypic fidelity, temporal resolution, and predictive scalability across genetically diverse patients.[17,77-79] However, LLMs, particularly those equipped with multi-modal embedding and transformer-based architectures,[80-82] are overcoming these limitations by parsing vast datasets that include single-cell RNA-seq, electrophysiological traces, and ion channel dynamics to generate high-resolution disease maps. These models are particularly valuable in predicting arrhythmogenic cardiomyopathy, long QT syndrome, and hypertrophic pathways by recognizing transcriptomic anomalies or delayed afterdepolarizations early in the iPSC-CM lifecycle.[83,84]

Diagnostic assistance has extended into automated interpretation of echocardiograms and coronary computed tomography angiography imaging, offering real-time triage support for acute coronary syndrome.[85] Clinically, the convergence of LLMs with real-time telemetry, EHR-derived biometrics, and wearable data streams is advancing early detection of ischemia, subclinical myocarditis, or mechanical desynchrony. In real-world applications, Japan's Keio University and the United States-based Stanford BioHub have documented significant improvements in outcomes using LLM-augmented surgical planning in congenital heart anomalies and heart failure risk scoring models.[85,86] These data pipelines to detect diastolic dysfunction signatures with a 30% improved lead-time over standard echo interpretations. These platforms employ supervised learning through attention-weighted tokenization of patient metadata, including age, genotype, medication history, and cardiac rhythm strips, resulting in temporally contextualized diagnostics. Natural language extraction from imaging reports and procedural notes also supports risk stratification in patients awaiting valve replacement or regenerative therapy.

In surgical contexts, LLMs are becoming indispensable to pre-operative planning for congenital heart disease and heart failure reconstruction. Here, iPSC-CM-derived functional readouts, integrated with 3D imaging and spatial transcriptomics, enable AI-generated surgical roadmaps.[87,88] Using reinforcement learning algorithms, platforms trained on surgical registries and intraoperative sensor data can recommend optimized graft placements, conduction system preservation strategies, or pharmacological adjuncts tailored to the patient's cellular profile.[89,90]

In addition to transcriptomic and electrophysiological modeling, LLMs have increasingly complement protein structure prediction tools, such as AlphaFold (DeepMind),[55,56] RoseTTAFold (Baker Lab),[56,73,79,80] and ESMFold (Meta AI),[67,68] to enable multi-layered diagnostics in iPSC-CM disease modeling. These AI-driven predictors decode 3D folding of cardiomyocyte-specific proteins, including titin (TTN),[91] myosin heavy chain 7 (MYH7),[92] sodium voltage-gated channel alpha subunit 5 (SCN5A),[93] and ryanodine receptors,[94] allowing structural annotation of patient-derived mutations and elucidating their pathogenic impact. For example, AlphaFold-enhanced variant analysis has been used to map missense-induced conformational changes in sarcomeric proteins, aligning well with LLM-predicted phenotypes, such as reduced contractility or altered calcium kinetics. This fusion of sequence-based and structure-based inference supports early diagnostics of inherited cardiomyopathies, including dilated or arrhythmogenic subtypes. Moreover, for Japan's Institute of Physical and Chemical Research and Germany's Max Planck Bioinformatics Lab, hybrid models integrating AlphaFold predictions with iPSC-CM drug testing platforms have identified altered drug-binding dynamics in mutated β1-adrenergic receptors, offering insight into individual therapeutic responsiveness.[95-98]

Clinically, this multilayered modeling assists surgical planning by flagging high-risk molecular defects before regenerative implantation, such as graft-host desmosome incompatibility in arrhythmia-prone myocardium. Thus, predictive diagnostics now extend beyond transcriptomes into the structural proteome, enabling cardiology to move from symptomatology to atomic-resolution risk stratification.[98,99]

Together, these advances represent a paradigm shift: from descriptive cardiomyocyte modeling to predictive, action-oriented diagnostics. LLMs not only enhance the resolution and interpretability of iPSC-CM-based disease simulation but also usher in an era where computational frameworks intersect with cardiomyocyte differentiation pathways, where neural networks model the heart across molecular and clinical scales to inform patient care. These implementations exemplify the synthesis of computational intelligence with biomolecular insight, elevating care delivery from reactive to proactive. Ultimately, this fusion of AI and cardiac physiology reflects a refined, forward-thinking pursuit—where innovation, integrity, and patient-centered design come together with clarity, elegance, and meaningful clinical impact.

Across the cited studies, LLM integration varies by both task and setting. For example, transcriptomic modeling by Li *et al.*[48] emphasizes mapping the regulatory pathway, while the study by Grafton *et al.*[100] focuses on detecting early cardiotoxicity. Furthermore, while BioGPT shows strength in knowledge synthesis, CardioGenAI demonstrates clinical-genetic alignment. These contrasts illustrate a spectrum from foundational modeling to translational precision, underscoring the importance of tailoring AI tools to specific regenerative goals.

3.3. Integration with EHRs and biomarkers

The fusion of LLMs with EHRs and biomarker datasets is accelerating the shift toward predictive, personalized cardiovascular care.[101,102] By analyzing structured and unstructured clinical data, including discharge summaries, imaging reports, laboratory trends, and physician notes, LLMs can extract subtle, temporally correlated patterns often missed by traditional models. For instance, leveraging expansive, open-access datasets such as Medical Information Mart for Intensive Care IV[103,104] allows LLMs to elegantly interweave structured and unstructured clinical information, unlocking nuanced, temporally aligned insights that illuminate early-stage cardiac dysfunction, including subtle diastolic anomalies in heart failure with preserved ejection fraction or asymptomatic ischemia in diabetic populations.[104,105]

For instance, in a recent meta-analysis by Zaka *et al.*,[116] machine-learning frameworks demonstrated superior risk stratification following percutaneous coronary intervention, outperforming conventional clinical models across multiple cohorts. Synthesizing data from global centers, including advanced cardiac units in Asia and North America, the study demonstrated that AI-driven, multimodal pipelines enhance predictive precision for major adverse cardiovascular events, setting a new benchmark for data-integrated, patient-tailored cardiology.[109] These findings align with parallel advancements reported by Tremamunno *et al.*[117] in the context of computed tomography-planned transcatheter aortic valve replacement, and by Chung *et al.*,[118] who highlighted the expanding role of LLMs in perioperative risk prediction and individualized prognostication.

By extracting nuanced clinical trajectories from EHRs, LLM-integrated platforms such as REALM and models trained on multimodal data are elevating precision in iPSC-CM research, enabling early phenotype-genotype matching, streamlining patient selection, and accelerating translational pathways from regenerative hypothesis to bedside impact.[110-127]

3.4. Therapeutic response prediction and drug screening

LLMs are increasingly applied to iPSC-CM drug screening, offering new tools for personalized cardiology and regenerative pharmacology. LLMs can be combined with phenotypic data from iPSC-CMs, such as calcium transients,[121,122] action potentials,[123,124] and contractility waveforms,[125,126] to simulate therapeutic responses across diverse, patient-derived cardiomyocytes. These models stratify compounds early, identifying effective therapies and flagging cardiotoxic risks before *in vivo* testing.[127]

Recent studies have highlighted the translational potential of AI-enhanced frameworks in cardiac safety pharmacology using human iPSC-CMs. For instance, Grafton *et al.*[100] used deep learning to detect cardiotoxicity with a higher sensitivity than immunofluorescence assays. Their models captured subtle shifts, such as QTc prolongation and mitochondrial changes, and linked them to known clinical cardiotoxic profiles.[100] These models identified subtle phenotypic changes, such as QTc prolongation and mitochondrial disruption, and linked them to known cardiotoxic profiles. Similarly, research in Frontiers in Pharmacology by Shim *et al.*[128] demonstrated that computational models integrating transcriptomic data and mechanistic simulations could predict individual-specific cardiotoxic responses to tyrosine kinase inhibitors. When validated against patient-derived iPSC-CMs, these predictions aligned with observed electrophysiological abnormalities, supporting the use of AI to anticipate drug-induced arrhythmias in genetically predisposed populations. Moreover, a study in Pharmaceutical Research[129] introduced a hybrid *in silico* platform that combined physiologically based pharmacokinetic and quantitative systems pharmacology models. By incorporating iPSC-CM-derived functional data, this platform accurately predicted the risk of systolic dysfunction in virtual patient cohorts receiving cardiotoxic chemotherapeutics, validating its utility against clinical endpoints. Collectively, these findings demonstrate how integrating AI and iPSC-CM platforms—especially with expanding capabilities of large-scale models—bridges predictive toxicology with regenerative medicine. The convergence of these technologies is paving the way for individualized drug safety screening and the rational design of therapeutics with minimized adverse cardiac effects.

Furthermore, LLMs' ability to consolidate multi-omic datasets—mining epigenomic, proteomic, and transcriptomic responses—enhances their utility in predicting adverse cardiac events with unprecedented temporal resolution.[127-132] *In silico* cardiotoxicity models, trained on extensive compound structure-toxicity literature, are now capable of flagging risks that might otherwise remain undetected in the early stages of screening.[133,134]

Real-world applications continue to emerge. For instance, Japan's collaboration between regenerative medicine institutes and AI developers has produced deep learning-assisted screenings of iPSC-CMs under anthracycline exposure, successfully predicting cardiotoxic thresholds in chemotherapy patients.[99] In parallel, an FDA-supported pilot study in the United States integrated LLM-driven safety models with iPSC-CMs from patients with complex arrhythmia syndromes, directly informing clinical decision-making by identifying therapeutic agents with both robust efficacy and minimal toxicity.[135]

Collectively, these advances not only shorten the bench-to-bedside timeline but also enhance patient safety by reducing the inherent trial-and-error burden in drug development. As precision therapies become increasingly molecularly targeted, LLMs are poised to propel cardiac regenerative medicine into a new era characterized by safer, more effective, and patient-responsive interventions.

3.5. Mechanistic and diagnostic integration through omics, CRISPRs, and NLP applications

LLMs are increasingly deployed to bridge mechanistic discovery and diagnostic translation in iPSC-CM research by systematically integrating CRISPR datasets, multi-omics layers, and clinical telemetry. These systems enable insight across three key domains: identification of transcriptional regulators, molecular mechanism mapping, and omics-enhanced diagnostics.[136-138]

For example, RoFormer-based and graph attention networks now facilitate high-resolution enhancer-promoter mapping, which has been validated in the context of Wnt and Notch signaling bifurcations. Likewise, transformer-based models, such as BioBERT and scGPT, have been integrated with ECG telemetry and transcriptomics to identify arrhythmic risk with lineage-specific precision,[118-125] successfully prioritizing core regulator genes—*TBX5*, *NKX2-5*, and *MEF2C*[66-77]—that define early cardiac lineage commitment. By mining large-scale literature corpora and chromatin interaction data, these models have also identified co-factors, including *GATA4*, *HAND2*, and *SIRT1*, which contribute to subtype specification and maturation.[141-143] Deep generative architectures, including RoFormer and graph attention networks, now enable high-resolution predictions of enhancer-promoter interactions, making them valuable tools for mapping mesoderm-to-cardiomyocyte transitions *in vitro*.[139,140]

Beyond regulatory insight, LLMs contribute to diagnostic augmentation by integrating multimodal omics data with patient telemetry and imaging. These models analyze ECG signals, cardiac CTs, and biomarker profiles to generate patient-specific readouts and multimodal disease signatures. This supports real-time triage and phenotype-genotype linkage in inherited cardiomyopathies, arrhythmia risk, and drug response profiling.[104-108,118-125] In particular, transformer models, including BioMedLM, LLaMA, and scGPT, demonstrate utility in combining transcriptomic features with electrophysiological telemetry from patient-derived iPSC-CMs to anticipate disease progression or treatment response.[118-122]

Recent translational efforts have extended this modeling to chromatin-level regulation. Japanese research teams, for instance, have integrated low-abundance enhancer data from patient-derived iPSC-CMs to reveal transcriptional noise patterns associated with dilated cardiomyopathy and impaired maturation signatures.[144] Concurrently, U.S.-based platforms have reconstructed mesoderm-to-cardiomyocyte developmental trajectories, uncovering regulatory bottlenecks in Wnt/β-catenin and Notch signaling cascades that influence fate decisions.[145,146] In rodents, long non-coding RNAs, such as Braveheart and histone demethylase-like lysine-specific demethylase 6A, have emerged as pivotal reprogramming regulators, suggesting that enhancer-focused LLMs may refine reprogramming fidelity at the chromatin interface.[141-143] LLMs analyze billions of molecular data points, enabling them to clarify complex biology and assist in mechanistic discovery, not just analytics. To streamline and avoid repetition, a consolidated table (Table 3) summarizes these

Table 3. Large language model functions across multi-omics integration, CRISPR insight, and diagnostic support

LLM function	Input data type	Models	Output/application
Gene editing target prioritization	CRISPR perturbation and scRNA-seq	BioBERT, scGPT	AI identified *TBX5*, *MEF2C*, and *NKX2-5* as core cardiac regulators, impacting the fate of iPSC-CM
Enhancer-promoter interaction mapping	Sequence, and epigenomic	Roformer, GAT (Graph Attention)	Predicted bifurcation nodes in Wnt/Notch pathways
Transcriptional co-factor discovery	Biomedical abstracts and protocol	BioMedLM	Revealed the influence of *GATA4*, *HAND2*, and *SIRT1* on subtype transitions
Lineage trajectory reconstruction	Chromatin maps, scRNA-seq, and ECG	Deep generative models	Modeled mesoderm-to-cardiomyocyte stages and stratified arrhythmia risk
Triage and diagnosis (biomarker inference)	ECT, CT, and telemetry	BiomedLM, LLaMA, and scGPT	Generated arrhythmia and cardiomyopathy risk profiles, predicted early fibrosis signal in cardiomyopathy
Variant interpretation	Multi-omics and phenotype	CardioGenAI	Linked gene variants to severity in inherited cardiac diseases

Abbreviations: AI: Artificial intelligence; CRISPR: Clustered regularly interspaced short palindromic repeats; CT: Computed tomography; ECG: Electrocardiogram; ECT: Electroconvulsive therapy; iPSC-CM: Induced pluripotent stem cell-derived cardiomyocytes; LLMs: Large language models; scRNA: Single-cell RNA.

integrated applications across mechanisms, models, and outputs.

3.6. Translational gaps and ethical risks

While the integration of LLMs into cardiovascular regenerative frameworks shows great promise, several systemic and technical limitations remain underexamined. These include generalizability across underrepresented populations, reproducibility of predictions in noisy or unstandardized datasets, and the interpretability of high-stakes clinical outputs, such as transplant decisions or differentiation outcomes.

A key concern is the validity of the cross-population model. Most LLMs in current use have been trained on data derived from high-income countries (HICs)—particularly the United States—European EHRs, biomedical literature, and clinical guidelines. As a result, model outputs may fail to generalize across populations with different genomic architectures, environmental stressors, and healthcare access patterns. For instance, LLMs trained exclusively on Western cardiac data have shown diminished sensitivity in detecting ischemic heart disease in Southeast Asian and rural African populations.[147-149] This bias not only impairs diagnostic accuracy but can also perpetuate disparities in regenerative therapy candidacy and outcome prediction.

Beyond data imbalance, biological noise and institutional heterogeneity also challenge reproducibility. iPSC-CM modeling involves variation across laboratory protocols, epigenetic memory effects, and differentiation batch variability.[150,151] These inconsistencies introduce latent confounders that can mislead LLM outputs, especially when working with small or institution-specific datasets. Furthermore, longitudinal datasets from low-resource regions remain scarce, limiting model calibration for predicting long-term outcomes, such as graft-host integration, ventricular remodeling, or sudden cardiac death.[152] Without multi-center validation pipelines and regionally calibrated metrics, LLMs risk producing brittle or misleading outputs under real-world biological and clinical complexity.

Ethical challenges compound these technical issues. LLMs trained on patient data raise privacy risks and call for enhanced frameworks for informed consent—particularly in iPSC-CM contexts where patient-derived cells are used for training predictive models.[156-158] In regenerative therapy, where interventions may be life-altering or irreversible, opacity of model logic is especially concerning. Clinicians must be able to interpret the reasons that a model recommends or predicts a given outcome; otherwise, reliance on black-box predictions in high-stakes decisions (e.g., transplant eligibility and cell graft rejection likelihood) could undermine patient safety and trust.

Finally, algorithmic bias remains a pressing concern. Models trained on skewed data distributions can unintentionally reinforce disparities in access to regenerative interventions, gender bias in diagnosis (e.g., underdiagnosis of women with microvascular disease), or triaging influenced by insurance status. These risks are magnified in low- and middle-income countries (LMIC) settings, where infrastructural gaps may be masked by generalized LLM outputs that do not account for resource constraints.

Moving forward, responsible deployment of LLMs in cardiovascular regenerative medicine demands global data equity, transparent architecture, and regulatory harmonization. Cross-continental consortia should be established to develop standardized, open-access cardiac datasets that incorporate genomic, imaging, and clinical data from underrepresented regions—including Japan, Indonesia, and countries in Latin America and Sub-Saharan Africa.[153-155] Only through such interdisciplinary, decentralized collaboration can AI-enabled regenerative medicine evolve in a way that is not only innovative, but also just, safe, and globally relevant.

3.7. Future directions and global equity

To unlock the full therapeutic scope of LLMs in cardiovascular regenerative medicine—particularly within iPSC-CM-based interventions—the next leap demands an infrastructure that is as globally inclusive as it is scientifically robust. LMICs, such as Indonesia, other ASEAN members, and regions across Sub-Saharan Africa, remain underrepresented in both clinical trial participation and regenerative medicine access. To correct this, scalable LLM-driven systems must be embedded into public health frameworks where analog records, inconsistent connectivity, and resource constraints are the norm. By integrating mobile diagnostics, point-of-care telemetry, and cloud-based EHR repositories, these systems can automate disease stratification, forecast trajectory shifts, and personalize post-transplant management even in decentralized care models.

The strategic development of federated learning ecosystems, in which anonymized cardiovascular datasets from diverse regions are collaboratively trained without breaching data sovereignty, ensures performance parity across ethnic, linguistic, and socioeconomic boundaries. Mobile LLM diagnostics, co-trained on electrophysiological data from iPSC-CM laboratories in Tokyo, Boston, and emerging hubs, have begun enhancing arrhythmia and ischemia detection in rural clinics. Crucially, these systems must be co-designed with local clinicians and patient communities to encode culturally relevant phenotypes and avoid epistemic asymmetries—thereby maximizing trust, usability, and precision.

Capacity-building remains essential. Regional training pipelines for clinicians and technologists alike must match the deployment of AI-regenerative tools in LMICs. Tele-education modules, academic exchange programs, and regional centers of excellence can catalyze local expertise and leadership. These efforts are beginning to materialize: academic-industry partnerships from Yogyakarta to Nairobi are already developing curriculum-integrated LLM training that supports both clinical interpretation and translational research design.

Yet amidst this momentum, a measured realism is necessary. While theoretical tools, such as real-time iPSC-CM protocol optimization, AI-assisted cryopreservation mapping, and graft-host compatibility prediction frameworks show enormous conceptual promise, many remain preclinical or unpublished. The peer-reviewed literature currently offers limited prototypes. However, early signals are emerging. Japan's Center for iPS Cell Research and Application, for instance, has piloted closed-loop AI platforms that fine-tune cardiomyocyte induction based on real-time metabolomic feedback. At Stanford, reinforcement-learning algorithms are being trained to simulate post-graft electrical integration using iPSC-CM-derived bio-signatures. Bioreactor-based cryopreservation mapping projects, aimed at predicting graft viability and post-thaw functionality, are also in conceptual testing. While these efforts remain in development, their presence marks the beginning of a tangible shift: from theoretical modeling to translational pipelines.

Hardware innovation must follow suit. Offline-compatible LLM interfaces and solar-powered diagnostic systems can mitigate bandwidth and electricity constraints in remote settings. Open-source software, policy-aligned governance, and shared trial infrastructures—backed by AI consortia, such as OpenAI and Hangzhou DeepseekAI—must underwrite this democratization.

In summary, the future of cardiovascular regenerative medicine rests not only in molecular innovation or computational elegance—but in the shared will to heal. When LLMs are built, deployed, and trusted across every corner of the healthcare spectrum, they cease being tools of privilege and become instruments of equity. This is the true legacy of an ethically coherent, biology-aligned, and human-centered cardiac future—where regenerative therapies reach every heart they are meant to save.

3.8. A comparative overview of LLMs in cardiovascular and regenerative contexts

Despite the rapid proliferation of LLMs in biomedical research, few studies have conducted systematic, domain-specific evaluations of their performance across regenerative and cardiovascular contexts. This absence of standardized benchmarking frameworks presents a notable gap in the translational landscape—particularly when considering the diverse architectural designs, training corpora, and deployment pipelines that shape each model's clinical relevance.

Emerging comparative studies have demonstrated that general-purpose LLMs, such as ChatGPT-4, excel in contextualizing clinical guidelines and summarizing literature with high fluency. However, they may underperform in multi-omics data integration due to a lack of domain-specific fine-tuning. In contrast, models such as BioGPT (Microsoft Research) and BioMedLM (Stanford

Table 4. Benchmarking key LLM and AI tools across cardiovascular and regenerative contexts

Name	Domain specificity	Primary input modality	Cardiovascular application	Key advantage
AlphaFold	Protein structure prediction	Amino acid sequences	Accurate modeling of cardiac proteins (e.g., sarcomere variants TTN, and MYH7)	High-resolution protein folding for CMs variant interpretation
AlphaMissense	Variant pathogenicity prediction	Gene variants	Interpret missense mutations, for example, in cardiomyopathy-related genes	Enables classifications of VUS in cardiac genomics using ClinVar-linked benchmarking
BioBERT	Biomedical NLP	Scholarly biomedical text	Named entity recognition and relation extraction in cardiology studies	Domain-tuned language understanding for gene-disease mining
BioGPT	Biomedical LLM (text generation and mining)	Biomedical text	Gene-disease annotation, literature summarization	High precision and recall in domain-specific NLP tasks
BioMedLM	Biomedical LLM	Text corpora of medical publications	Competitive QA performance on medical exams (~57–69%), QA systems in medical informatics, preliminary cardiovascular insights	Strong domain-specific NLP for QA
Cardiogen AI	Cardiac genomics ML	Genomic variant profiles	Predicting disease phenotype severity in monogenic CVDs	Superior variant-to-outcome interpretation in cardiology
ChatGPT-4	General-purpose LLM	Broad text corpora and multimodal inputs	Clinical guideline interpretation, literature synthesis, and preclinical planning	broad fluency and multi-step reasoning capability
Chemputer	Automated synthesis	Chemical synthesis pipelines	*In silico* synthesis for cardiac-regenerative compound generation	Automated drug-generation workflows tied to target biology
ClinVar	Variant database	raw clinical variant records	Reference database for variant pathogenicity annotation	Standard resource for variant interpretation benchmarking
DeepChem	Drug modeling library	*In silico* molecular prediction	Toxicity screening of compounds in cardiac assays	Efficient compound efficacy and toxicity modeling tools
DeepSeek-R1/Med	General-purpose LLM (China-owned)	Bilingual reasoning tasks	Potential use in the Chinese cardiovascular research context	Scalable, open-source model with strong multilingual NLP, rivaling GPT4
Ensembl Genome Browser	Genomic data platform	Genomic and transcriptomic query	Identifying regulatory regions and variants relevant to the iPSC-CM pipeline	Centralized gene/variant annotation hub
ESMFold	Structure-prediction and evolutionary LLM	Protein sequence	Efficient structure prediction aiding cardiac variant annotation	Fast and scalable folding predictions alternative to Alphafold
GEO	Public gene expression repository	Transcriptomic microarray and RNA-seq	Data source for cardiac gene expression variation and iPSC-CM training	Large-scale expression datasets for CM modeling
GROK	open source LLM (xAI)	Text reasoning	Emerging general reasoning tasks, limited iPSC-CM application yet	Early-stage reasoning capabilities in open models
HuggingFace Transformers	Model library and fine-tuning hub	NLP/ML frameworks	Used to fine-tune BioBERT/BioGPT/REALM for cardiac-specific tasks	Ecosystem support with model sharing and fine-tuning infrastructure
JAX	ML computation framework	Neural network training	Training LLMs or multimodal models for iPSC-CM omics interpretation	High-performance, accelerated neural architecture support
PyTorch	ML Framework	Deep-learning neural modeling	Foundation of variant annotation and regression models in cardiac biology	Large community and ecosystem for model development
REALM	Retrieval-augmented LLM	Document retrieval and text modeling	EHR mining and real-time guideline retrieval in cardiology workflows	Efficient integration of large-text archives with LLM query mechanisms

(Cont'd)

Table 4. ***(Continued)***

Name	Domain specificity	Primary input modality	Cardiovascular application	Key advantage
RoseTTAfold (Baker Lab)	Protein folding tool	Protein sequence and distance embedding	Structural prediction for variant evaluation in cell modeling platforms	Efficient three-track network for accurate fold prediction
scFoundation	Single-cell foundation model	Single-cell transcriptomics	Embeddings used for drug response prediction and lineage inference in iPSC-CMs	Outperforming baseline models in cell population mapping
scGPT	Single-cell generative transformer	Multi-omic cell atlas modeling	Predictive modeling of cell fate trajectories and disease phenotypes in cardiomyocyte differentiation	Scalable multi-scale modeling across millions of single cells
TensorFlow	Deep learning framework	Neural network model construction	Used in image, sequence, and time-series modeling within cardiac AI	Widely supported, with high interoperability across platforms

Abbreviations: AI: Artificial intelligence; CMs: Cardiomyocytes; CVDs: Cardiovascular diseases; EHRs: Electronic health records; iPSC-CM: Induced pluripotent stem cell-derived cardiomyocytes; LLM: Large language model; ML: Machine learning; MYH7: Myosin heavy chain 7; NLP: Natural language processing; QA: Quality assurance; TTN: Titin; VUS: Variance of uncertain significance.

CRFM) have been specifically optimized for biomedical corpora, demonstrating superior performance in gene-disease association tasks and biomolecular annotation, particularly in cardiomyopathy and arrhythmia literature mining.[59,63,159] A comparative summary of key models and their cardiovascular applications is presented in Table 4 to illustrate these performance distinctions across architectural and functional axes.

DeepSeekMed, a bilingual biomedical LLM trained on Chinese and English datasets, has outperformed ChatGPT and BioBERT in cross-lingual phenotype extraction and EHR-based cardiovascular risk scoring, making it especially promising for LMIC and multilingual health systems.[160-162] Similarly, CardioGenAI, a machine learning framework developed by BGI Genomics, has demonstrated high predictive accuracy in linking genetic variants with phenotype severity in monogenic CVDs, offering a focused advantage in genotype-phenotype interpretation relevant to iPSC-CM disease modeling.[163,164]

In structural biology and regenerative cardiology, AlphaFold2, RoseTTAFold, and ESMFold represent next-generation protein structure prediction tools that have outpaced prior algorithms in predicting conformational changes in sarcomeric proteins, such as MYH7, TTN, and SCN5A, the key targets in inherited cardiomyopathies. These models have been successfully integrated into iPSC-CM variant modeling platforms to anticipate functional disruptions before *in vitro* phenotyping.[165-167]

Several frameworks—such as BioGPT, BioMedLM, and scGPT—are particularly optimized for biomedical corpora and high-dimensional omics data, enabling improved performance in gene-disease association mining, transcriptomic trajectory modeling, and drug response prediction. Meanwhile, foundational platforms, such as JAX, TensorFlow, and PyTorch, remain essential for training custom cardiac models from raw datasets, offering flexibility in integrating imaging, text, and bio signal data streams.

Despite these advances, there remains no unified evaluation framework to assess LLM performance across core regenerative tasks, such as differentiation protocol optimization, cardiotoxicity modeling, or graft-host interaction simulation. Each model tends to be benchmarked independently, often using proprietary metrics, narrow data types, or single-institution validation sets, limiting clinical transferability. The lack of harmonized performance indicators, common cardiac data benchmarks, and population-representative validation pipelines—especially in iPSC-CM contexts—continues to hinder reproducibility and deployment scalability in clinical-grade environments.

Therefore, this review serves to highlight not only the current strengths and domain-specific niches of various LLMs in regenerative cardiology but also the urgent need for cross-institutional benchmarking consortia. Such efforts should incorporate:

(i) Standardized cardiac datasets (e.g., iPSC-CM electrophysiology, omics, and CRISPR perturbations)
(ii) Multimodal integration tasks (e.g., combining text, imaging, and gene expression), and
(iii) Regionally calibrated validation protocols (especially across LMIC and diverse genetic populations).

The current review functions not only as a descriptive summary but also as a framework proposal—a scaffold on which future empirical benchmarking protocols and clinical-grade LLM applications in cardiovascular regenerative medicine can be systematically developed.

Table 5. Proposed LLM-iPSC-CM evaluation framework

LLM	Output types	Benchmark task	Suggested metric	Application in iPSC-CM
AlphaFold	Protein 3D structure prediction	Structural mutation mapping	RMSD/TM-score	Sarcomeric protein modeling for inherited cardiac diseases
AlphaMissense	Pathogenicity classification	Missense variant pathogenicity scoring	ClinVar concordance/PPV	Predicting clinical impact of sarcomeric mutations
BioBERT	Biomedical NER/relation extraction	Disease-gene-drug linkage mining	Precision/recall/F1 score	Mapping arrhythmia genes and drug interactions
BioGPT	Biomedical relation extraction	Gene-disease association mining	Precision/recall/F1 score	Prioritizing cardiomyopathy targets
BioMedLM	Literature summarization	Biomedical passage summarization	ROUGE-L/BERTScore	Rapid review of regenerative medicine papers
Cardiogen AI	Variant-phenotype linking	SNP-to-clinical outcome prediction	AUC/Matthews correlation coefficient	Personalized risk stratification using iPSC-CM
ChatGPT-4	Text generation/Q&A	Clinical Guideline interpretation	BLEU/ROUGE/expert rating	Patient education, therapeutic summarization
Chemputer	Chemical reaction planning	iPSC-CM-compatible media prediction	Reaction yield prediction accuracy	Media optimization for differentiation/stability
ClinVar	Clinical variant database	Variant validation for disease relevance	Overlap with patient variant sets	Validation of iPSC-CM patient-derived mutations
DeepChem	Molecular graph prediction	Drug-toxicity prediction	ROC/AUC/ sensitivity-specificity	Cardiotoxicity modeling via iPSC-CM
DeepSeek-R1/Med	Bilingual phenotype extraction	EHR-to-concept mapping in a multilingual setting	Exact match/recall	LMIC-compatible phenotype extraction
Ensembl Genome Browser	Gene annotation/visualization platform	iPSC-CM-related gene discovery	Annotation depth/retrieval accuracy	Regulatory target mining for cardiac differentiation
ESMFold	End-to-end structure generation	Cell lineage reconstruction	Trajectory concordance/ PAGA metrics	Maturation-state-specific folding (e.g., fetal vs. adult CM)
GEO	Omics data repository	Benchmarking gene expression in iPSC-CMs	Expression match score/ TPM fold-change	Model training dataset for transcriptomic-based prediction
GROK	Explainable AI output	Interpretability of iPSC-CM risk models	SHAP/LIME agreement with expert annotations	Enhancing transparency in regenerative risk models
HuggingFace Transformers	Model zoo and training framework	Deployment of biomedical transformer-based LLMs	Adaptability/API integration score	Hosting custom cardiac LLMs like fine-tuned BioGPT
JAX/PyTorch/ TensorFlow	Backend frameworks (for training custom models)	Custom LLM implementation/ fine-tuning	Neural FLOPs/time to convergence/accuracy	Infrastructure layer for cardiac LLM pipelines
REALM	Document retrieval	Omics data-linked literature navigation	Recall@10/NDCG	Evidence mining for protocol optimization
RoseTTAfold (Baker Lab)	Protein-protein interaction prediction	Binding site inference	Interface RMSD/DockQ	Drug-target screening via iPSC-CM
scFoundation	Single-cell foundation model	Generalization across cardiac single-cell datasets	Silhouette score/batch effect reduction	Cross-cohort prediction in cardiac developmental states
scGPT	Single-cell trajectory generation	Cell lineage reconstruction	Trajectory concordance/ PAGA metrics	iPSC-to-cardiomyocyte fate modeling

Abbreviations: 3D: Three-dimensional; API: Application programming interface; AUC: Area under the curve; BERT: Bidirectional encoder representations from transformers; BLEU: Bilingual evaluation understudy; CM: Cardiomyocytes; EHR: Electronic health record; FLOPs: Floating point operations per second; iPSC-CM: Induced pluripotent stem cell-derived cardiomyocytes; LIME: Local interpretable model-agnostic explanations; LLM: Large language model; LMIC: Low- and middle-income countries; NDCG: Normalized discounted cumulative gain; NER: Named entity recognition; PAGA: Partition-based graph abstraction; PPV: Positive predictive value; Q&A: Question and answer; RMSD: Root mean square deviation; ROC: Receiver operating characteristic; ROUGE: Recall-oriented understudy for Gisting evaluation; SHAP: Shapley additive explanation; SNP: Single-nucleotide polymorphism; TM: Template modeling; TPM: Transcripts per million.

4. Conclusion

4.1. Synthesizing the path forward

LLMs have emerged not as passive computational tools but as cognitive collaborators in the evolution of cardiovascular regenerative medicine. Their symbiosis with iPSC-CM technologies has redefined the possibility of bridging molecular depth with clinical foresight, transforming static data into dynamic, patient-specific insight. From decoding transcriptomic vulnerabilities to simulating drug responses and unmasking hidden cardiac signaling cascades, LLMs elevate regenerative cardiology from a discipline of promise to a praxis of precision.

Through the lens of iPSC-derived cardiomyocytes, LLMs do not merely predict outcomes—they actively co-shape them. Whether parsing the hidden linguistics of cardiac electrical signals or annotating the silent language of mutated sarcomeric genes, these models act as translators between biology's complexity and medicine's intent. Institutions in Japan, the United States, and others are already weaving LLMs into clinical pipelines, illustrating a future where human intuition and machine intelligence are harmoniously aligned in rhythm and resolution.

4.2. Limitations of this review

While this review presents a comprehensive outlook on the applications of LLMs in iPSC-CM research, it is not without limitations. First, the field is rapidly evolving, and novel models—particularly multimodal foundation models integrating text, images, and omics—emerge at a pace that risks outdating current interpretations. The current analysis also leans heavily on literature and infrastructural models from HICs, which may not fully account for the logistical and technological constraints present in LMICs. Challenges, such as limited access to high-throughput iPSC-CM platforms, fragmented EHRs, and low local computational capacity, may hinder the real-world use of LLM-based tools in these regions. While ethically aligned LLM deployment is emphasized, this review cannot substitute for the legal, clinical, and sociotechnical audits necessary before practical implementation in research or care. It does not offer a validated benchmarking pipeline, nor does it provide quantitative evaluations of model accuracy. Major technical limitations include the lack of validation across genetically diverse populations, insufficient quantification of uncertainties in LLM-driven predictions, and nascent regulatory frameworks for AI-based regenerative therapies. These gaps impede reproducibility and hinder clinical translation, particularly in the context of LMICs. Furthermore, the interpretability of transformer-based predictions remains a black-box challenge, demanding post-hoc explainability layers before regulatory approval. Nonetheless, by framing both upstream and downstream application nodes, this review offers a conceptual scaffold to examine ethical, biological, and translational fault lines in LLM-guided regenerative medicine.

Other important limitations remain, such as ethical risks (e.g., erroneous decisions made based on automated predictions), issues of bias in biological data, the need for rigorous regulation and clinical validation, and institutional resistance to AI integration.

This review should be seen as both a map and a mirror—a reflection of current achievements and a roadmap for future empirical validation. Although fundamentally theoretical, it is grounded in real-world implementations of LLMs and structured to highlight performance differentials, practical gaps, and future benchmarks across AI platforms in cardiovascular regenerative medicine.

4.3. Future directions

The next frontier lies in the intentional integration of LLMs with wet-lab protocols and clinical trials. This includes dynamic LLM-based systems that adjust differentiation protocols in real-time based on omics feedback from iPSC-CM cultures, AI-assisted cryopreservation mapping to ensure graft integrity, or predictive frameworks for long-term graft-host interactions post-transplantation. In addition, LLMs must transition from being interpreters of known science to generators of new hypotheses, supporting regenerative surgeons and electrophysiologists in exploring novel frontiers of cardiac identity, within the constraints of current interpretability and validation frameworks.

Globally, the emphasis must shift toward algorithmic equity, particularly through the development of federated and multilingual models so that iPSC-CM-based therapeutics do not remain a privilege of academic elites but a birthright of every human heart—regardless of geography, gross domestic product, or genetic background. A cross-continental commons for cardiac data, rooted in transparency and cultural humility, could democratize access and imagination.

To catalyze translational acceleration, we propose a modular evaluation framework that cross-references LLMs by output type (e.g., protocol optimization, variant annotation, and predictive modeling), validation method (e.g., wet-lab cross-check and patient data alignment), and regulatory stage (research-only, preclinical, or investigational use) (Table 5).

To summarize, LLMs are currently supporting cardiovascular regenerative research by interpreting omics data, optimizing iPSC-CM differentiation protocols, and simulating clinical outcomes. Their most immediate promise lies in transforming static cardiac datasets into real-time, patient-specific insights—especially when integrated with iPSC-CM platforms and longitudinal health data. Critical gaps remain in model reproducibility, equity in data representation, and clinical translation—particularly in low-resource settings and non-Western genomic contexts. To move from theoretical potential to clinical practice, the field must prioritize federated learning, interdisciplinary education, ethical model design, and globally inclusive infrastructures that decentralize innovation and democratize access.

4.4. Final reflection on AI in the service of empathy

The integration of LLMs into regenerative cardiology offers a powerful means to enhance clinical precision and therapeutic foresight. Rather than replacing human judgment, these systems are designed to augment clinicians' capabilities—supporting nuanced interpretation and broader reach. As medicine enters an era shaped by both molecular insight and computational intelligence, a balanced and ethically guided approach is essential.

The success of LLMs in cardiovascular applications will depend not only on their technical performance but also on their alignment with human values. Empathy, transparency, and accountability must remain central, ensuring that these models are developed and deployed in ways that respect patient agency and clinical nuance.

Acknowledgments

None.

Funding

None.

Conflict of interest

Tadahisa Sugiura is an Editorial Board Member of this journal and Guest Editor of this special issue, but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Separately, other authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Author contributions

Conceptualization: Dhienda C. Shahannaz, Tadahisa Sugiura
Writing–original draft: Dhienda C. Shahannaz, Tadahisa Sugiura
Writing–review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Di Cesare M, Perel P, Taylor S, *et al*. The heart of the world. *Global Heart*. 2024;19(1):11.

   doi: 10.5334/gh.1288

2. World Health Organization: WHO. *Cardiovascular Diseases (CVDs)*; 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [Last accessed on 2025 May 20].

3. Hussain MA, Mamun AA, Peters SA, Woodward M, Huxley RR. The burden of cardiovascular disease attributable to major modifiable risk factors in Indonesia. *J Epidemiol*. 2016;26(10):515-521.

   doi: 10.2188/jea.je20150178

4. Pratama DA, Anisah M, Marlianda SA. Real-Time changes of heart rhythm using MATLAB. *Int J Res Vocat Stud*. 2023;3(2):53-59.

   doi: 10.53893/ijrvocas.v3i2.202

5. Maddox KEJ, Elkind MSV, Aparicio HJ, *et al*. Forecasting the burden of cardiovascular disease and stroke in the United States through 2050-Prevalence of Risk Factors and Disease: A Presidential advisory from the American Heart Association. *Circulation*. 2024;150(4):e65-e88.

   doi: 10.1161/cir.0000000000001256

6. Zhu M, Jin W, He W, Zhang L. The incidence, mortality and disease burden of cardiovascular diseases in China: A comparative study with the United States and Japan based on the GBD 2019 time trend analysis. *Front Cardiovasc Med*. 2024;11:1408487.

   doi: 10.3389/fcvm.2024.1408487

7. Singh R, Chandi SK, Sran S, *et al*. Emerging therapeutic strategies in cardiovascular diseases. *Cureus*. 2024;16:e64388.

   doi: 10.7759/cureus.64388

8. Ullah A, Kumar M, Sayyar M, *et al*. Revolutionizing cardiac care: A comprehensive narrative review of cardiac rehabilitation and the evolution of cardiovascular medicine. *Cureus*. 2023;15:e46469.

   doi: 10.7759/cureus.46469

9. World Heart Federation. *CVD Prevention What we Do World Heart Federation*. World Heart Federation; 2024. Available from: https://world-heart-federation.org/what-we-do/prevention/#:~:text=The%20majority%20of%20deaths%20due,regular%20exercise%20and%20avoiding%20tobacco [Last accessed on 2025 May 20].

10. Addissouky TA, Sayed IETE, Ali MMA, *et al.* Shaping the future of cardiac wellness: Exploring revolutionary approaches in disease management and prevention. *J Clin Cardiol*. 2024;5(1):6-29.

    doi: 10.33696/cardiology.5.048

11. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126(4):663-676.

    doi: 10.1016/j.cell.2006.07.024

12. Sugiura T, Shahannaz DC, Ferrell BE, Yoshida T. Advancements in cardiac regenerative therapy: Scalable human iPSC-derived cardiomyocyte differentiation and maturation. *Glob Transl Med*. 2025;4(1):1-15.

    doi: 10.36922/gtm.5745

13. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Bus Horiz*. 2018;62(1):15-25.

    doi: 10.1016/j.bushor.2018.08.004

14. Wang Z, Chu Z, Doan TV, Ni S, Yang M, Zhang W. History, Development, and Principles of Large Language Models: An Introductory Survey. *AI and Ethics*; 2024.

    doi: 10.1007/s43681-024-00583-7

15. Olawade DB, Aderinto N, Olatunji G, Kokori E, David-Olawade AC, Hadi M. Advancements and applications of Artificial Intelligence in cardiology: Current trends and future prospects. *J Med Surg Public Health*. 2024;3:100109.

    doi: 10.1016/j.glmedi.2024.100109

16. Tolu-Akinnawo OZ, Ezekwueme F, Omolayo O, Batheja S, Awoyemi T. Advancements in artificial intelligence in noninvasive cardiac imaging: A comprehensive review. *Clin Cardiol*. 2025;48(1):e70087.

    doi: 10.1002/clc.70087

17. Kasartzian DI, Tsiampalis T. Transforming cardiovascular risk prediction: A review of machine learning and artificial intelligence innovations. *Life*. 2025;15(1):94.

    doi: 10.3390/life15010094

18. Leivaditis V, Beltsios E, Papatriantafyllou A, *et al.* Artificial intelligence in Cardiac surgery: Transforming outcomes and shaping the future. *Clin Pract*. 2025;15(1):17.

    doi: 10.3390/clinpract15010017

19. Wangsa K, Karim S, Gide E, Elkhodr M. A systematic review and comprehensive analysis of pioneering AI chatbot models from education to healthcare: ChatGPT, Bard, Llama, Ernie and Grok. *Future Internet*. 2024;16(7):219.

    doi: 10.3390/fi16070219

20. Iqbal U, Tanweer A, Rahmanti AR, Greenfield D, Lee LTJ, Li YCJ. Impact of large language model (ChatGPT) in healthcare: An umbrella review and evidence synthesis. *J Biomed Sci*. 2025;32(1):45.

    doi: 10.1186/s12929-025-01131-z

21. Chen H, Zhang S, Zhang L, *et al*. Multi role ChatGPT framework for transforming medical data analysis. *Sci Rep*. 2024;14(1):13930.

    doi: 10.1038/s41598-024-64585-5

22. Neha F, Bhati D, Shukla DK, Amiruzzaman M. ChatGPT: Transforming healthcare with AI. *AI*. 2024;5(4):2618-2650.

    doi: 10.3390/ai5040126

23. Liu S, Wright AP, Patterson BL, *et al*. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *J Am Med Inform Assoc*. 2023;30(7):1237-1245.

    doi: 10.1093/jamia/ocad072

24. Salihu A, Meier D, Noirclerc N, *et al.* A study of ChatGPT in facilitating Heart Team decisions on severe aortic stenosis. *EuroIntervention*. 2024;20(8):e496-e503.

    doi: 10.4244/eij-d-23-00643

25. Ahmed SK, Hussein S, Essa RA. The role of ChatGPT in cardiothoracic surgery. *Indian J Thorac Cardiovasc Surg*. 2023;39(5):562-563.

    doi: 10.1007/s12055-023-01568-7

26. Amacher SA, Arpagaus A, Sahmer C, *et al*. Prediction of outcomes after cardiac arrest by a generative artificial intelligence model. *Resusc Plus*. 2024;18:100587.

    doi: 10.1016/j.resplu.2024.100587

27. Clark SC. Can ChatGPT transform cardiac surgery and heart transplantation? *J Cardiothorac Surg*. 2024;19(1):108.

    doi: 10.1186/s13019-024-02541-0

28. Sharun K, Banu SA, Pawde AM, *et al.* ChatGPT and artificial hallucinations in stem cell research: Assessing the accuracy of generated references – a preliminary study. *Ann Med Surg*. 2023;85(10):5275-5278.

    doi: 10.1097/ms9.0000000000001228

29. Yap L, Adusumalli S, Lim S, *et al.* Spatiotemporal transcriptomics of human cardiovascular progenitors in pig hearts identifies MIDKINE as a positive regulator of neovascularization. *Research Square*. 2025.

    doi: 10.21203/rs.3.rs-5837914/v1

30. Murphy SA, Chen EZ, Tung L, Boheler KR, Kwon C. Maturing heart muscle cells: Mechanisms and transcriptomic insights. *Semin Cell Dev Biol*. 2021;119:49-60.

    doi: 10.1016/j.semcdb.2021.04.019

31. Hao X, Zhao J, Rodriguez-Wallberg KA. Comprehensive atlas of mitochondrial distribution and dynamics during oocyte maturation in mouse models. *Biomark Res*. 2024;12(1):125.

doi: 10.1186/s40364-024-00672-z

32. Talman V, Teppo J, Pöhö P, *et al.* Molecular atlas of postnatal mouse heart development. *J Am Heart Assoc*. 2018;7(20):e010378.

    doi: 10.1161/jaha.118.010378

33. Hayat R. Dynamics of metabolism and regulation of epigenetics during cardiomyocytes maturation. *Cell Biol Int*. 2022;47(1):30-40.

    doi: 10.1002/cbin.11931

34. Huang L, Wang Q, Gu S, Cao N. Integrated metabolic and epigenetic mechanisms in cardiomyocyte proliferation. *J Mol Cell Cardiol*. 2023;181:79-88.

    doi: 10.1016/j.yjmcc.2023.06.002

35. Rommel C, Hein L. Four dimensions of the cardiac myocyte epigenome: From fetal to adult heart. *Curr Cardiol Rep.* 2020;22(5):26.

    doi: 10.1007/s11886-020-01280-7

36. Zamora-Dorta M, Laine-Menéndez S, Abia D, *et al.* Time-resolved mitochondrial screen identifies regulatory components of oxidative metabolism. *EMBO Rep.* 2025;26:3045-3074.

    doi: 10.1038/s44319-025-00459-9

37. Liu B, Chang H, Yang D, *et al.* A deep learning framework assisted echocardiography with diagnosis, lesion localization, phenogrouping heterogeneous disease, and anomaly detection. *Sci Rep.* 2023;13:3.

    doi: 10.1038/s41598-022-27211-w

38. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and machine learning for heart failure survival analysis. *Stud Health Technol Inform*. 2015;216:40-44.

    doi: 10.3233/978-1-61499-564-7-40

39. Pan J, Lee S, Cheligeer C, *et al.* Integrating large language models with human expertise for disease detection in electronic health records. *Comput Biol Med*. 2025;191:110161.

    doi: 10.1016/j.compbiomed.2025.110161

40. Zhu Y, Huang R, Wu Z, *et al.* Deep learning-based predictive identification of neural stem cell differentiation. *Nat Commun*. 2021;12:2614.

    doi: 10.1038/s41467-021-22758-0

41. Pickard J, Choi MA, Oliven N, *et al. Bioinformatics Retrieval Augmentation Data (BRAD) Digital Assistant*. New York: arXiv Cornell University; 2024.

    doi: 10.48550/arxiv.2409.02864

42. Sapunov G. *Deep Learning with JAX.* United States: Simon and Schuster; 2024.

43. Hao M, Wei L, Yang F, *et al*. Current opinions on large cellular models. *Quant Biol*. 2024;12(4):433-443.

    doi: 10.1002/qub2.65

44. Hochstadt A, Barbhaiya C, Aizer A, *et al.* Performance of a protein language model for variant annotation in cardiac disease. *J Am Heart Assoc*. 2024;13(20):e036921.

    doi: 10.1161/jaha.124.036921

45. Llm-Jp, Aizawa A, Aramaki E, *et al. LLM-JP: A Cross-Organizational Project for the Research and Development of Fully Open Japanese LLMs*. New York: arXiv Cornell University; 2024.

    doi: 10.48550/arxiv.2407.03963

46. Cui H, Wang C, Maan H, *et al*. scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods*. 2024;21(8):1470-1480.

    doi: 10.1038/s41592-024-02201-0

47. Hao M, Gong J, Zeng X, *et al. Large Scale Foundation Model on Single-Cell Transcriptomics. bioRxiv* New York: Cold Spring Harbor Laboratory; 2023.

    doi: 10.1101/2023.05.29.542705

48. Li Y, Mamouei M, Khorshidi GS, *et al. Hi-BEHRT: Hierarchical Transformer-based Model for Accurate Prediction of Clinical Events Using Multimodal Longitudinal Electronic Health Records*. New York: arXiv Cornell University; 2021.

    doi: 10.48550/arxiv.2106.11360

49. Ning Z, Jiang X, Huang H, *et al.* Machine learning integration of multimodal data identifies key features of circulating NT-proBNP in people without cardiovascular diseases. *Sci Rep.* 2025;15(1):12015.

    doi: 10.1038/s41598-025-96689-x

50. Neyazi M, Bremer JP, Knorr MS, *et al*. Deep learning-based NT-proBNP prediction from the ECG for risk assessment in the community. *Clin Chem Lab Med*. 2023;62(4):740-752.

    doi: 10.1515/cclm-2023-0743

51. Gunčar G, Kukar M, Smole T, *et al*. Differentiating viral and bacterial infections: A machine learning model based on routine blood test values. *Heliyon*. 2024;10(8):e29372.

    doi: 10.1016/j.heliyon.2024.e29372

52. Maxwell YL. In HFREF, *AI Shows Promise for Predicting Beta Blocker Response*; 2021. Available from: https://www.tctmd.com/news/hfref-ai-shows-promise-predicting-beta-blocker-response [Last accessed on 2025 May 20].

53. Kyro GW, Martin MT, Watt ED, Batista VS. *CardioGenAI: A Machine Learning-Based Framework for Re-Engineering Drugs for Reduced HERG Liability*. New York: arXiv Cornell University; 2024.

    doi: 10.48550/arxiv.2403.07632

54. Chiu CE, Pinto AL, Chowdhury RA, Christensen K, Varela M. *Characterisation of Anti-Arrhythmic Drug Effects on Cardiac Electrophysiology using Physics-Informed Neural*

Networks. New York: arXiv Cornell University; 2024.

doi: 10.1109/isbi56570.2024.10635234

55. Ouyang JF, Chothani S, Rackham OJL. Deep learning models will shape the future of stem cell research. *Stem Cell Reports*. 2023;18(1):6-12.

    doi: 10.1016/j.stemcr.2022.11.007

56. Sen S, Hoff S, Bonomi M. BPS2025-Improving small-molecule docking using AlphaFold models with bAIes. *Biophys J*. 2025;124(3):224a.

57. Wysocki O, Zhou Z, O'Regan P, *et al*. *Transformers and the Representation of Biomedical Background Knowledge*. New York: arXiv Cornell University; 2022.

    doi: 10.48550/arxiv.2202.02432

58. Lee J, Yoon W, Kim S, *et al.* BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2019;36(4):1234-1240.

    doi: 10.1093/bioinformatics/btz682

59. Luo R, Sun L, Xia Y, *et al*. BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Brief Bioinform*. 2022;23(6):bbac409.

    doi: 10.1093/bib/bbac409

60. Bolton E, Venigalla A, Yasunaga M, *et al. BioMedLM: A 2.7B Parameter Language Model Trained on Biomedical Text*; 2024. Available from: https://arxiv.org/abs/2403.18421v1 [Last accessed on 2025 May 20].

61. Cronin L. *Chemputer and Chemputation - a Universal Chemical Compound Synthesis Machine*; 2024. Available from: https://arxiv.org/abs/2408.09171v2 [Last accessed on 2025 May 20].

62. Peng C, Yang X, Chen A, *et al*. A study of generative large language model for medical research and healthcare. *NPJ Digit Med*. 2023;6(1):210.

    doi: 10.1038/s41746-023-00958-w

63. Chen JH, Tseng YJ. A general optimization protocol for molecular property prediction using a deep learning network. *Brief Bioinform*. 2021;23(1):bbab367.

64. Nosrati H, Nosrati M. Artificial intelligence in regenerative medicine: Applications and implications. *Biomimetics (Basel)*. 2023;8(5):442.

    doi: 10.3390/biomimetics8050442

65. Herson J, Krummenacker M, Spaulding A, O'Maille P, Karp PD. The genome explorer genome browser. *mSystems*. 2024;9(7):e0026724.

    doi: 10.1128/msystems.00267-24

66. Forero DA, Bonilla DA, González-Giraldo Y, Patrinos GP. An overview of key online resources for human genomics: A powerful and open toolbox for *in silico* research. *Brief Funct Genomics*. 2024;23(6):754-764.

    doi: 10.1093/bfgp/elae029

67. Nijkamp E, Ruffolo JA, Weinstein EN, Naik N, Madani A. ProGen2: Exploring the boundaries of protein language models. *Cell Syst*. 2023;14(11):968-978.e3.

    doi: 10.1016/j.cels.2023.10.002

68. Kulkarni P, Porter L, Chou TF, *et al.* Evolving concepts of the protein universe. *iScience*. 2025;28(3):112012.

    doi: 10.1016/j.isci.2025.112012

69. Clough E, Barrett T, Wilhite SE, *et al*. NCBI GEO: Archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res*. 2023;52(D1):D138-D144.

    doi: 10.1093/nar/gkad965

70. Stanescu L, Dinu G. TensorFlow vs. PyTorch in classifying medical images – preliminary results. In: *2022 26th International Conference on System Theory, Control and Computing (ICSTCC)*; 2023. p. 448-453.

    doi: 10.1109/icstcc59206.2023.10308472

71. Novac OC, Chirodea MC, Novac CM, *et al*. Analysis of the application efficiency of TensorFlow and PyTorch in convolutional neural network. *Sensors (Basel)*. 2022;22(22):8872.

    doi: 10.3390/s22228872

72. Sugiura T, Shahannaz DC, Ferrell BE. Current status of cardiac regenerative therapy using induced pluripotent stem cells. *Int J Mol Sci*. 2024;25(11):5772.

    doi: 10.3390/ijms25115772

73. Sugiura T, Nawaz S, Shahannaz DC, Ferrell BE, Yoshida T. From injury to repair: The therapeutic potential of induced pluripotent stem cells in heart failure. *Regen Med Rep*. 2025;2(1):22-30.

    doi: 10.4103/regenmed.regenmed-d-25-00002

74. Shahannaz DC, Sugiura T, Ferrell BE. Enhancing mitochondrial maturation in IPSC-DerivedCardiomyocytes: Strategies for metabolic optimization. *BioChem*. 2025;5(3):23.

    doi: 10.3390/biochem5030023

75. Shi Y, Yang J, Nai C, *et al*. *Language-Enhanced representation learning for Single-Cell transcriptomics*; 2025. Available from: https://arxiv.org/abs/2503.09427 [Last accessed on 2025 May 20].

76. Wang Y, Chen X, Zheng Z, *et al.* scGREAT: Transformer-based deep-language model for gene regulatory network inference from single-cell transcriptomics. *iScience*. 2024;27(4):109352.

    doi: 10.1016/j.isci.2024.109352

77. Mao Y, Mi Y, Liu P, Zhang M, Liu H, Gao Y. *SCAgent: Universal Single-Cell Annotation via a LLM agent*; 2025. Available from: https://arxiv.org/abs/2504.04698 [Last

accessed on 2025 May 20].

78. Collins RL, Glessner JT, Porcu E, *et al.* A cross-disorder dosage sensitivity map of the human genome. *Cell*. 2022;185(16):3041-3055.e25.

    doi: 10.1016/j.cell.2022.06.036

79. Zhu Y, Ren C, Xie S, *et al. REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models*; 2024. Available from: https://arxiv.org/abs/2402.07016v1 [Last accessed on 2025 May 20].

80. Krishna R, Wang J, Ahern W, *et al*. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science*. 2024;384(6693):eadl2528.

    doi: 10.1126/science.adl2528

81. Bunne C, Roohani Y, Rosen Y, *et al.* How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell*. 2024;187(25):7045-7063.

    doi: 10.1016/j.cell.2024.11.015

82. Keshri R, Detraux D, Phal A, *et al*. Next-generation direct reprogramming. *Front Cell Dev Biol*. 2024;12:1343106.

    doi: 10.3389/fcell.2024.1343106

83. Lam WY, Au SCL. From ChatGPT to DeepSeek: Potential uses of artificial intelligence in early symptom recognition for stroke care. *J Acute Dis*. 2025;14(1):6.

    doi: 10.4103/jad.jad_16_25

84. Lavrov AV, Varenikov GG, Skoblov MY. Genome scale analysis of pathogenic variants targetable for single base editing. *BMC Med Genom*. 2020;13(Suppl 8):80.

    doi: 10.1186/s12920-020-00735-8

85. Keio University. *AI Model Developed by Brigham Researchers Could Help Screen for Heart Defect*. Keio University; 2023. Available from: https://www.keio.ac.jp/en/press-releases/files/2023/11/7/231107-1.pdf [Last accessed on 2025 May [27].

86. Miura K, Yagi R, Miyama H, *et al*. Deep learning-based model detects atrial septal defects from electrocardiography: A cross-sectional multicenter hospital-based study. *EClinicalMedicine*. 2023;63:102141.

    doi: 10.1016/j.eclinm.2023.102141

87. Batteux C, Haidar MA, Bonnet D. 3D-printed models for surgical planning in complex congenital heart diseases: A systematic review. *Front Pediatr*. 2019;7:23.

    doi: 10.3389/fped.2019.00023

88. Sørensen TS, Beerbaum P, Mosegaard J, *et al*. Virtual cardiotomy based on 3-D MRI for preoperative planning in congenital heart disease. *Pediatr Radiol*. 2008; 38(12):1314-1322.

    doi: 10.1007/s00247-008-1032-5

89. Staffa SJ, Zurakowski D. A basic machine learning primer for surgical research in congenital heart disease. *World J Pediatr Congen Heart Surg*. 2025;16:571-577.

    doi: 10.1177/21501351251335643

90. Holt DB, El-Bokl A, Stromberg D, Taylor MD. Role of artificial intelligence in congenital heart disease and interventions. *J Soc Cardiovasc Angiogr Interv*. 2025;4(3):102567.

    doi: 10.1016/j.jscai.2025.102567

91. Micheu MM, Rosca AM. Patient-specific induced pluripotent stem cells as "disease-in-a-dish" models for inherited cardiomyopathies and channelopathies – 15 years of research. *World J Stem Cells*. 2021;13(4):281-303.

    doi: 10.4252/wjsc.v13.i4.281

92. Funakoshi S, Yoshida Y. Recent progress of iPSC technology in cardiac diseases. *Arch Toxicol*. 2021;95(12):3633-3650.

    doi: 10.1007/s00204-021-03172-3

93. Hong L, Zhang M, Ly OT, *et al*. Human induced pluripotent stem cell-derived atrial cardiomyocytes carrying an SCN5A mutation identify nitric oxide signaling as a mediator of atrial fibrillation. *Stem Cell Reports*. 2021;16(6):1542-1554.

    doi: 10.1016/j.stemcr.2021.04.019

94. Guven O, DeMirci H. *Structural Analysis and Docking Studies of FK506-Binding Protein 1A*. *bioRxiv*. New York: Cold Spring Harbor Laboratory; 2025.

    doi: 10.1101/2025.05.22.655516

95. *RIKEN-Max Planck Joint Research Center for Systems Chemical Biology*. RIKEN. Available from: https://www.riken.jp/en/collab/research/riken_mpg/ [Last accessed on 2025 May 20].

96. Xu X, Kaindl J, Clark MJ, *et al.* Binding pathway determines norepinephrine selectivity for the human β1AR over β2AR. *Cell Res*. 2020;31(5):569-579.

    doi: 10.1038/s41422-020-00424-2

97. Numata G, Otsu Y, Nakamura S, *et al. In vivo* effects of Cardiomyocyte-Specific Beta-1 blockade on afterload- and frequency-dependent cardiac performance. *Am J Physiol Heart Circ Physiol*. 2025;328:H543-H549.

    doi: 10.1152/ajpheart.00795.2024

98. Cantwell CD, Mohamied Y, Tzortzis KN, *et al*. Rethinking multiscale cardiac electrophysiology with machine learning and predictive modelling. *Comput Biol Med*. 2018;104:339-351.

    doi: 10.1016/j.compbiomed.2018.10.015

99. Fatkin D, Calkins H, Elliott P, James CA, Peters S, Kovacic JC. Contemporary and future approaches to precision medicine in inherited cardiomyopathies. *J Am Coll Cardiol*. 2021;77(20):2551-2572.

    doi: 10.1016/j.jacc.2020.12.072

100. Grafton F, Ho J, Ranjbarvaziri S, *et al*. Deep learning detects cardiotoxicity in a high-content screen with induced pluripotent stem cell-derived cardiomyocytes. *eLife*.

2021;10:e68714.

doi: 10.7554/elife.68714

101. Li L, Zhou J, Gao Z, *et al*. *A Scoping Review of Using Large Language Models (LLMs) to Investigate Electronic Health Records (EHRs)*; 2024. Available from: https://arxiv.org/abs/2405.03066v2 [Last accessed on 2025 May 20].

102. Tsai M, Chen K, Chen P. Harnessing electronic health records and artificial intelligence for enhanced cardiovascular Risk Prediction: A Comprehensive review. *J Am Heart Assoc*. 2025;14:e036946.

doi: 10.1161/jaha.124.036946

103. Kucharska-Newton AM, Loop MS, Bullo M, *et al*. Use of troponins in the classification of myocardial infarction from electronic health records. The Atherosclerosis Risk in Communities (ARIC) Study. *Int J Cardiol*. 2021;348:152-156.

doi: 10.1016/j.ijcard.2021.12.022

104. Johnson AEW, Bulgarelli L, Shen L, *et al*. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1.

doi: 10.1038/s41597-022-01899-x

105. Nagamine T, Gillette B, Kahoun J, Burghaus R, Lippert J, Saxena M. Data-driven identification of heart failure disease states and progression pathways using electronic health records. *Sci Rep*. 2022;12(1):7871.

doi: 10.1038/s41598-022-22398-4

106. Singhal P, Tan ALM, Drivas TG, Johnson KB, Ritchie MD, Beaulieu-Jones BK. Opportunities and challenges for biomarker discovery using electronic health record data. *Trends Mol Med*. 2023;29(9):765-776.

doi: 10.1016/j.molmed.2023.06.006

107. Bhasuran B, Manoharan S, Iyyappan OR, Murugesan G, Prabahar A, Raja K. Large language models and genomics for summarizing the role of microRNA in regulating mRNA expression. *Biomedicines*. 2024;12(7):1535.

doi: 10.3390/biomedicines12071535

108. Du X, Wang Y, Zhou Z, *et al*. *Generative Large Language Models in Electronic Health Records for Patient Care Since 2023: A Systematic Review. medRxiv*. New York: (Cold Spring Harbor Laboratory); 2024.

doi: 10.1101/2024.08.11.24311828

109. Wells QS, Gupta DK, Smith JG, *et al*. Accelerating biomarker discovery through electronic health records, automated biobanking, and proteomics. *J Am Coll Cardiol*. 2019;73(17):2195-2205.

doi: 10.1016/j.jacc.2019.01.074

110. Zhang J, Chen Z, Ma M, He Y. Soluble ST2 in coronary artery disease: Clinical biomarkers and treatment guidance. *Front Cardiovasc Med*. 2022;9:924461.

doi: 10.3389/fcvm.2022.924461

111. Pelletier AR, Ramirez J, Adam I, *et al*. *Explainable Biomedical Hypothesis Generation via Retrieval Augmented Generation enabled Large Language Models*; 2024. Available from: https://arxiv.org/abs/2407.12888v1 [Last accessed on 2025 May 20].

112. Raza MZ, Xu J, Lim T, *et al*. *LLM-TA: An LLM-Enhanced Thematic Analysis Pipeline for Transcripts from Parents of Children with Congenital Heart Disease*; 2025. Available from: https://arxiv.org/abs/2502.01620v1 [Last accessed on 2025 May 20].

113. Kong F, Stocker S, Choi PS, Ma M, Ennis DB, Marsden A. *SDF4CHD: Generative Modeling of Cardiac Anatomies with Congenital Heart Defects*; 2023. Available from: https://arxiv.org/abs/2311.00332v2 [Last accessed on 2025 May 20].

114. Simon ST, Mandair D, Tiwari P, Rosenberg MA. Prediction of drug-induced long QT Syndrome using machine learning applied to harmonized electronic health record data. *J Cardiovasc Pharmacol Ther*. 2021;26(4):335-340.

doi: 10.1177/1074248421995348

115. Zhang H, Tarabanis C, Jethani N, *et al*. QTNet: Predicting drug-induced QT prolongation with artificial intelligence-enabled electrocardiograms. *JACC Clin Electrophysiol*. 2024;10(5):956-966.

doi: 10.1016/j.jacep.2024.01.022

116. Zaka A, Mutahar D, Gorcilov J, *et al*. Machine-learning approaches for risk prediction after percutaneous coronary intervention: A systematic review and meta-analysis. *Eur Heart J Digit Health*. 2024;6(1):23-44.

doi: 10.1093/ehjdh/ztae074

117. Tremamunno G, Vecsey-Nagy M, Schoepf UJ, *et al*. Artificial intelligence improves prediction of major adverse cardiovascular events in patients undergoing transcatheter aortic valve replacement planning CT. *Acad Radiol*. 2024;32(2):702-711.

doi: 10.1016/j.acra.2024.09.046

118. Chung P, Fong CT, Walters AM, Aghaeepour N, Yetisgen M, O'Reilly-Shah VN. Large language model capabilities in perioperative risk prediction and prognostication. *JAMA Surg*. 2024;159(8):928.

doi: 10.1001/jamasurg.2024.1621

119. Kazaki N, Hattori K, Shota H, *et al*. *Building a Large Japanese Web Corpus for Large Language Models*. Available from: https://arxiv.org/html/2404.17733v1 [Last accessed on 2025 May 20].

120. Singhal K, Azizi S, Tu T, *et al*. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180.

doi: 10.1038/s41586-023-06291-2

121. Yang H, Stebbeds W, Francis J, *et al*. Deriving waveform

parameters from calcium transients in human iPSC-derived cardiomyocytes to predict cardiac activity with machine learning. *Stem Cell Reports*. 2022;17(3):556-568.

doi: 10.1016/j.stemcr.2022.01.009

122. Yang H, Obrezanova O, Pointon A, *et al.* Prediction of inotropic effect based on calcium transients in human iPSC-derived cardiomyocytes and machine learning. *Toxicol Appl Pharmacol*. 2022;459:116342.

    doi: 10.1016/j.taap.2022.116342

123. Martin CH, Oved A, Chowdhury RA, *et al.* EP-*PINNs: Cardiac Electrophysiology Characterisation using Physics-Informed Neural Networks*; 2021. Available from: https://arxiv.org/abs/2112.07703v1 [Last accessed on 2025 May 20].

124. Árpádffy-Lovas T, Nagy N. ActionPytential: An open source tool for analyzing and visualizing cardiac action potential data. *Heliyon*. 2023;9(3):e14440.

    doi: 10.1016/j.heliyon.2023.e14440

125. Hoang P, Jacquir S, Lemus S, Ma Z. Quantification of contractile dynamic complexities exhibited by human stem Cell-Derived cardiomyocytes using nonlinear dimensional analysis. *Sci Rep*. 2019;9(1):14714.

    doi: 10.1038/s41598-019-51197-7

126. Naghavi E, Wang H, Fan L, *et al.* Rapid estimation of left ventricular contractility with a physics-informed neural network inverse modeling approach. *Artif Intell Med*. 2024;157:102995.

    doi: 10.1016/j.artmed.2024.102995

127. Lieber A, Kiem HP. Prospects and challenges of *in vivo* hematopoietic stem cell genome editing for hemoglobinopathies. *Mol Ther*. 2023;31(10):2823-2825.

    doi: 10.1016/j.ymthe.2023.09.006

128. Shim JV, Xiong Y, Dhanan P, *et al.* Predicting individual-specific cardiotoxicity responses induced by tyrosine kinase inhibitors. *Front Pharmacol*. 2023;14:1158222.

    doi: 10.3389/fphar.2023.1158222

129. Sang L, Zhou Z, Luo S, *et al.* An *in silico* platform to predict cardiotoxicity risk of anti-tumor drug combination with hiPSC-CMs based *in vitro* study. *Pharm Res*. 2023;41(2):247-262.

    doi: 10.1007/s11095-023-03644-4

130. Nam Y, Kim J, Jung SH, *et al.* Harnessing Artificial intelligence in multimodal Omics Data Integration: Paving the path for the next frontier in precision medicine. *Annu Rev Biomed Data Sci*. 2024;7(1):225-250.

    doi: 10.1146/annurev-biodatasci-102523-103801

131. Khera R, Asnani AH, Krive J, *et al.* Artificial intelligence to enhance precision medicine in cardio-oncology: A scientific statement from the American Heart Association. *Circ Genom Precis Med*. 2025;18:e000097.

    doi: 10.1161/hcg.0000000000000097

132. Pinero SL, Li X, Zhang J, *et al. Omics-Based Computational Approaches for Biomarker Identification, Prediction, and Treatment of Long COVID*. *MedRxIV*; 2025.

    doi: 10.1101/2025.04.01.25324942

133. Pramudito MA, Fuadah YN, Qauli AI, Marcellinus A, Lim KM. Explainable artificial intelligence (XAI) to find optimal in-silico biomarkers for cardiac drug toxicity evaluation. *Sci Rep*. 2024;14(1):24045.

    doi: 10.1038/s41598-024-71169-w

134. Simon ST, Trinkley KE, Malone DC, Rosenberg MA. Interpretable machine learning prediction of drug-induced QT prolongation: Electronic health record analysis. *J Med Internet Res*. 2022;24(12):e42163.

    doi: 10.2196/42163

135. Research C for DEA. *Cardiomyocytes for Mechanistic Cardiovascular Safety Liabilities. U.S. Food And Drug Administration*; 2019. Available from: https://www.fda.gov/drugs/news-events-human-drugs/cardiomyocytes-mechanistic-cardiovascular-safety-liabilities [Last accessed on 2025 May 20].

136. Tan WLW, Seow WQ, Zhang A, *et al.* Current and future perspectives of single-cell multi-omics technologies in cardiovascular research. *Nat Cardiovasc Res*. 2023;2(1):20-34.

    doi: 10.1038/s44161-022-00205-7

137. Yang C, Jin Y, Yin Y. Integration of single-cell transcriptome and chromatin accessibility and its application on tumor investigation. *Life Med*. 2024;3(2):lnae015.

    doi: 10.1093/lifemedi/lnae015

138. Pierce SE, Granja JM, Greenleaf WJ. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. *Nat Commun*. 2021;12(1):2969.

    doi: 10.1038/s41467-021-23213-w

139. Song M, Wen J, Kuang Y, Xie M. EPI-RMDL: Prediction of enhancer-promoter interactions based on ROFormer mechanism and deep learning. In: *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; 2024. p. 357-362.

    doi: 10.1109/bibm62325.2024.10821931

140. Zhang T, Zhao X, Sun H, Gao B, Liu X. GATv2EPI: Predicting enhancer-promoter interactions with a dynamic graph attention network. *Genes*. 2024;15(12):1511.

    doi: 10.3390/genes15121511

141. Klattenhoff CA, Scheuermann JC, Surface LE, *et al.* Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. 2013;152(3):570-583.

    doi: 10.1016/j.cell.2013.01.003

142. Xue Z, Hennelly S, Doyle B, *et al.* A G-Rich Motif in the lncRNA Braveheart interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. *Mol Cell.* 2016;64(1):37-50.

doi: 10.1016/j.molcel.2016.08.010

143. Maurya SS, Yang W, Zhang Q, *et al. KDM6A Knockout in Human iPSCs Alters the Genome-wide Histone Methylation Profile at Active and Poised Enhancers, Activating Expression of Ectoderm Gene Expression Pathways. bioRxiv*. New York: Cold Spring Harbor Laboratory; 2021.

doi: 10.1101/2021.03.09.434633

144. Hayashi H, Ko T, Dai Z, *et al.* TRAITER: Transformer-guided diagnosis and prognosis of heart failure using cell nuclear morphology and DNA damage marker. *Bioinformatics*. 2024;40(11):btae610.

doi: 10.1093/bioinformatics/btae610

145. Cyganek L, Tiburcy M, Sekeres K, *et al*. Deep phenotyping of human induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes. *JCI Insight*. 2018;3(12):e99941.

doi: 10.1172/jci.insight.99941

146. Yang D, Gomez-Garcia J, Funakoshi S, *et al.* Modeling human multi-lineage heart field development with pluripotent stem cells. *Cell Stem Cell*. 2022;29(9):1382-1401.e8.

doi: 10.1016/j.stem.2022.08.007

147. Rahman S, Jiang LY, Gabriel S, Aphinyanaphongs Y, Oermann EK, Chunara R. *Generalization in Healthcare AI: Evaluation of a Clinical Large Language Model*; 2024. Available from: https://arxiv.org/abs/2402.10965v2 [Last accessed on 2025 May 20].

148. Quer G, Topol EJ. The potential for large language models to transform cardiovascular medicine. *Lancet Digit Health*. 2024;6(10):e767-e771.

doi: 10.1016/s2589-7500(24)00151-1

149. Gendler M, Nadkarni GN, Sudri K, *et al. Large Language Models in Cardiology: A Systematic Review*. *medRxiv*. New York: Cold Spring Harbor Laboratory; 2024.

doi: 10.1101/2024.09.01.24312887

150. Volpato V, Webber C. Addressing variability in iPSC-derived models of human disease: Guidelines to promote reproducibility. *Dis Models Mech*. 2020;13(1):dmm042317.

doi: 10.1242/dmm.042317

151. Kim K, Doi A, Wen B, *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature*. 2010;467(7313):285-290.

doi: 10.1038/nature09342

152. Chai B, Efstathiou C, Yue H, Draviam VM. Opportunities and challenges for deep learning in cell dynamics research. *Trends Cell Biol*. 2023;34(11):955-967.

doi: 10.1016/j.tcb.2023.10.010

153. Li Z, Brittan M, Mills NL. A multimodal OMIcS framework to empower target discovery for cardiovascular regeneration. *Cardiovasc Drugs Ther*. 2023;38(2):223-236.

doi: 10.1007/s10557-023-07484-7

154. Reitz CJ, Kuzmanov U, Gramolini AO. Multi-omic analyses and network biology in cardiovascular disease. *Proteomics*. 2023;23(21-22):e202200289.

doi: 10.1002/pmic.202200289

155. Alemu R, Sharew NT, Arsano YY, *et al.* Multi-omics approaches for understanding gene-environment interactions in noncommunicable diseases: Techniques, translation, and equity issues. *Hum Genomics*. 2025;19(1):8.

doi: 10.1186/s40246-025-00718-9

156. Isasi R, Bentzen HB, Fabbri M, *et al*. Dynamic governance: A new era for consent for stem cell research. *Stem Cell Reports*. 2024;19(9):1233-1241.

doi: 10.1016/j.stemcr.2024.07.006

157. Orzechowski M, Schochow M, Kühl M, Steger F. Content and method of information for participants in clinical studies with Induced Pluripotent stem cells (IPSCS). *Front Cell Dev Biol*. 2021;9:627816.

doi: 10.3389/fcell.2021.627816

158. Borziak K, Parvanova I, Finkelstein J. ReMeDy: A platform for integrating and sharing published stem cell research data with a focus on iPSC trials. *Database*. 2021;2021:baab038.

doi: 10.1093/database/baab038

159. Tian S, Jin Q, Yeganova L, *et al*. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Brief Bioinform*. 2023;25(1):bbad493.

doi: 10.1093/bib/bbad493

160. Al-Janabi O, Alyasiri OM, Jebur EA, Nafl SM. Evaluating AI language models in news retrieval: A comparative study of ChatGPT-Plus and DeepSeek (R1). *InfoTech Spectr Iraqi J Data Sci*. 2024;2(2):14-20.

doi: 10.51173/ijds.v2i2.33

161. Wu J, Wang Z, Qin Y. Performance of Deepseek-R1 and ChatGPT-4O on the Chinese national medical licensing examination: A comparative study. *J Med Syst*. 2025;49(1):74.

doi: 10.1007/s10916-025-02213-z

162. Lin S, Duan Y, Zhou T, Liu X, Wang J. EHMQA-GPT: A knowledge Augmented large language model for personalized elderly health management. *Information*. 2025;16(6):467.

doi: 10.3390/info16060467

163. Kyro GW, Martin MT, Watt ED, Batista VS. CardioGenAI: A machine learning-based framework for re-engineering drugs for reduced hERG liability. *J Cheminform*. 2025;17(1):30.

doi: 10.1186/s13321-025-00976-8

164. Jumper J, Evans R, Pritzel A, *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589.

doi: 10.1038/s41586-021-03819-2

165. Baek M, DiMaio F, Anishchenko I, *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021;373(6557):871-876.

doi: 10.1126/science.abj8754

166. Fang Y, Jiang Y, Wei L, *et al.* DeepProSite: Structure-aware protein binding site prediction using ESMFold and pretrained language model. *Bioinformatics*. 2023;39(12):btad718.

doi: 10.1093/bioinformatics/btad718

167. Lin Z, Akin H, Rao R, *et al*. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. 2023;379(6637):1123-1130.

doi: 10.1126/science.ade2574