

REVIEW ARTICLE

Large language models as integrative
intelligence for multimodal cardiovascular
decision supportPredrag Mitrovic^{1,2*}¹Cardiology Clinic, University Clinical Center of Serbia, Belgrade, Serbia²Faculty of Medicine, University of Belgrade, Belgrade, Serbia

Abstract

Background: In this narrative review, we examine large language models (LLMs) as an emerging component of cardiovascular artificial intelligence and propose the concept of integrative intelligence as a physician-supervised orchestration framework rather than a new model architecture. Current cardiovascular evidence remains heterogeneous and includes text-based LLMs, multimodal foundation or vision-language models, and dedicated modality-specific algorithms for electrocardiography (ECG), echocardiography, computed tomography (CT), cardiovascular magnetic resonance imaging (CMR), and physiological signals. Accordingly, conventional text-based LLMs should not be assumed to interpret raw cardiovascular waveforms or images directly. Instead, their most plausible near-term role is to synthesize clinical text, structured electronic health record data, biomarkers, retrieved evidence, and validated outputs from specialized analytical systems. **Methods:** We reviewed applications in acute coronary syndromes, heart failure, arrhythmias, valvular disease, cardio-oncology, documentation, and decision support, while distinguishing clinically evaluated applications from proof-of-concept and proposed future uses. **Results:** Available randomized and prospective evidence is still limited, and reported benefits are task-dependent; improvements in diagnostic reasoning, workflow, or intermediate decision-support measures should not be interpreted as established reductions in major cardiovascular events, readmissions, or mortality. We therefore emphasize data acquisition and multimodal fusion, external and prospective validation, calibration, omission errors, hallucinations, prompt sensitivity, automation bias, model drift, subgroup and out-of-distribution performance, evidence provenance, retrieval-augmented generation failure modes, cybersecurity, regulation, governance, and professional accountability. **Conclusion:** Integrative intelligence is presented as a testable translational framework for combining specialized AI, multimodal models, evidence retrieval, and human clinical judgment. Its clinical value will depend on rigorous multicenter validation, transparent auditability, and preservation of physician responsibility. **Relevance for Patients:** This review clarifies the emerging role of LLMs in cardiovascular medicine and proposes a physician-supervised framework for safe multimodal clinical integration.

Keywords: Large language models; Cardiovascular medicine; Artificial intelligence; Multimodal data integration; Clinical decision support; Electronic health records

***Corresponding author:**Predrag Mitrovic
(predragm@email.com)

Citation: Mitrovic P. Large language models as integrative intelligence for multimodal cardiovascular decision support. *J Clin Transl Res*. doi: 10.36922/JCTR026290071

Received: July 18, 2026**Revised:** August 11, 2026**Accepted:** August 24, 2026**Published online:** September 15, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0), which permits all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Cardiovascular diseases (CVDs) remain the leading cause of morbidity and mortality worldwide despite substantial advances in diagnostic technologies, evidence-based therapies, and preventive strategies. Contemporary cardiovascular care generates an unprecedented volume of heterogeneous clinical data, including electronic health records (EHRs), electrocardiograms (ECGs), echocardiographic images, laboratory biomarkers, medical imaging, clinical reports, and guideline-based recommendations. Although each data source provides valuable information, their effective integration into a coherent clinical decision remains a major challenge in everyday practice.

Artificial intelligence (AI) has rapidly emerged as a promising solution for addressing this complexity. During the past decade, machine learning and deep learning algorithms have demonstrated remarkable performance in cardiovascular imaging, automated ECG interpretation, risk prediction, and outcome modeling. Nevertheless, most existing AI systems have been developed to solve narrowly defined tasks, such as image classification, segmentation, or prediction of a single clinical outcome. Consequently, these models often operate in isolation and lack the ability to synthesize information from multiple clinical modalities in a manner similar to human clinical reasoning.

The recent development of large language models (LLMs) represents a major paradigm shift in medical artificial intelligence. Unlike conventional machine learning algorithms, LLMs are capable of understanding natural language, summarizing complex medical information, interpreting clinical documentation, generating structured reports, and integrating both structured and unstructured data into context-aware responses. Initially developed for natural language processing, LLMs are increasingly being explored as clinical reasoning systems capable of supporting diagnosis, therapeutic decision-making, patient communication, scientific writing, and medical education.

Cardiovascular medicine is particularly well suited for the implementation of LLM-based technologies because clinical decision-making inherently depends on multimodal data integration. For example, the diagnosis and management of acute coronary syndrome require simultaneous interpretation of symptoms, serial ECGs, cardiac biomarkers, echocardiographic findings, coronary imaging, previous medical history, and current guideline recommendations. Similarly, patients with heart failure, cardiac arrhythmias, or valvular heart disease require continuous synthesis of longitudinal clinical information

collected from multiple diagnostic platforms. Rather than replacing specialist interpretation, LLMs have the potential to function as an intelligent interface capable of organizing, contextualizing, and prioritizing complex cardiovascular information for clinician review.¹⁻³

Recent investigations have demonstrated encouraging performance of LLMs in several cardiovascular applications, including automated report generation, extraction of clinically relevant information from EHRs, interpretation of guideline-based recommendations, clinical question answering, patient education, and support for cardiovascular research. Furthermore, the emergence of multimodal foundation models capable of integrating text, medical images, physiological signals, and laboratory data suggests that future AI systems may evolve beyond isolated diagnostic tools toward comprehensive clinical decision-support platforms.^{2,4-67}

Despite these promising developments, important challenges continue to limit routine clinical implementation. Hallucinated outputs, variable factual accuracy, limited explainability, potential algorithmic bias, privacy concerns, regulatory requirements, and the need for prospective clinical validation remain significant barriers before LLMs can be safely integrated into cardiovascular practice. Consequently, current evidence supports the use of LLMs as clinician-support systems rather than autonomous decision makers, emphasizing the importance of maintaining human oversight in all critical clinical decisions.^{1,3,6}

This narrative review presents a contemporary overview of the emerging role of LLMs in cardiovascular medicine while proposing a conceptual framework in which these models serve as integrative intelligence rather than standalone conversational systems. We discuss how LLMs can bridge echocardiography, electrocardiography, cardiac biomarkers, and EHRs to facilitate multimodal clinical reasoning, improve workflow efficiency, and support precision cardiovascular medicine. Finally, we examine current limitations, ethical considerations, and future research directions required for the safe translation of LLM-based technologies into routine cardiovascular care.

Terminology used throughout this review is deliberately specific: “LLM” refers primarily to text-based large language models; “multimodal foundation model” or “vision-language model” denotes models explicitly trained to process more than one input modality; “dedicated cardiovascular AI” refers to modality-specific analytical algorithms; and “integrative intelligence” refers to the physician-supervised orchestration framework proposed here rather than to a distinct model architecture.

2. Literature search strategy

This article was designed as a narrative review rather than a systematic review or meta-analysis. A targeted literature search was conducted in PubMed/MEDLINE, supplemented by backward citation tracking from relevant reviews and primary studies and by targeted searches of journal and publisher websites. The search was updated through August 2026 and focused primarily on literature published from 2018 onward, while earlier landmark studies were retained when necessary to describe the evolution of artificial intelligence in cardiovascular medicine.

Search terms were used individually and in combinations and included “large language model,” “LLM,” “generative artificial intelligence,” “foundation model,” “multimodal artificial intelligence,” “vision-language model,” “cardiovascular,” “cardiology,” “electrocardiography,” “ECG,” “echocardiography,” “cardiac imaging,” “computed tomography,” “cardiovascular magnetic resonance,” “electronic health record,” “clinical decision support,” “retrieval-augmented generation,” “hallucination,” “validation,” “governance,” and “regulation.” Reference lists of eligible articles were additionally screened to identify relevant studies not captured by the initial search.

Priority was given to peer-reviewed original investigations, prospective or externally validated studies, systematic or scoping reviews, consensus or reporting frameworks, major cardiovascular guidelines, and authoritative regulatory or governance documents. Studies were included when they addressed LLMs, multimodal foundation models, dedicated cardiovascular AI systems relevant to multimodal integration, or clinical orchestration and decision-support workflows. Articles were excluded when they were unrelated to cardiovascular medicine, lacked sufficient methodological information, represented non-substantive commentary without direct relevance to the conceptual framework, or duplicated evidence already represented by a more complete or updated publication.

Because the evidence base is rapidly evolving and heterogeneous in model architecture, input modality, clinical task, validation design, and reported outcomes, no formal meta-analysis was attempted. Evidence was therefore synthesized narratively, with particular attention to distinguishing clinically evaluated applications from proof-of-concept studies and proposed or future uses. Where clinical outcome data were unavailable, potential benefits are described as anticipated or hypothesis-generating rather than established effects.

3. Evolution of artificial intelligence in cardiovascular medicine

The evolution of AI in cardiovascular medicine has been characterized by a progressive transition from rule-based expert systems to increasingly sophisticated machine learning (ML), deep learning (DL), foundation models, and, most recently, LLMs. This evolution reflects not only advances in computational capabilities but also the growing availability of digital cardiovascular data generated through routine clinical practice. Modern cardiology has become one of the most data-intensive medical specialties, relying on information derived from electrocardiography (ECG), echocardiography, cardiac computed tomography (CT), cardiovascular magnetic resonance imaging (CMR), laboratory biomarkers, wearable devices, and EHRs. Consequently, AI has emerged as a key technology for extracting clinically meaningful information from these heterogeneous data sources.^{9,10}

The earliest applications of AI in cardiology were primarily based on expert systems that relied on manually encoded clinical rules. Although these systems successfully supported limited diagnostic tasks, they lacked adaptability and were unable to learn from continuously expanding clinical datasets. The introduction of machine learning algorithms represented a substantial improvement by enabling automated identification of statistical patterns associated with cardiovascular diseases. Supervised learning models such as logistic regression, random forests, support vector machines, and gradient boosting algorithms demonstrated improved performance in cardiovascular risk prediction, mortality estimation, and outcome prediction following acute myocardial infarction and heart failure.^{9,11}

The subsequent emergence of deep learning further transformed cardiovascular diagnostics. Convolutional neural networks (CNNs) enabled automated interpretation of cardiovascular imaging, particularly echocardiography, coronary CT angiography, and cardiovascular magnetic resonance imaging, while recurrent neural networks and transformer-based architectures substantially improved ECG interpretation. Numerous studies demonstrated that deep learning algorithms achieved expert-level performance in left ventricular function assessment, chamber segmentation, valvular disease classification, arrhythmia detection, and prediction of future cardiovascular events directly from ECG signals. Importantly, these systems reduced observer variability while improving workflow efficiency and diagnostic reproducibility.^{9,10}

Despite these remarkable achievements, most conventional AI models remained task-specific. Algorithms developed for ECG analysis could not directly interpret echocardiographic findings, whereas imaging models were generally unable to incorporate laboratory biomarkers, clinical history, or physician notes. Consequently, clinicians continued to perform the cognitive integration of multiple diagnostic modalities, limiting the full translational potential of AI in routine cardiovascular care.^{10,12}

Recent advances in foundation models have introduced a fundamentally different paradigm. Rather than being optimized for a single clinical task, foundation models are pretrained on massive multimodal datasets and subsequently adapted to diverse downstream applications through transfer learning. In cardiovascular medicine, these models increasingly combine medical images, physiological waveforms, structured clinical variables, and free-text documentation, enabling more generalized clinical reasoning than traditional AI algorithms.^{2,13}

Large language models (LLMs) represent the latest stage of this technological evolution. Initially designed for natural language understanding and text generation, LLMs have rapidly expanded beyond conversational applications to become knowledge-based reasoning systems capable of interpreting clinical documentation, summarizing medical evidence, extracting relevant information from EHRs, generating structured reports, and supporting guideline-based clinical decision-making. Unlike previous generations of AI that primarily analyzed isolated datasets, LLMs can synthesize heterogeneous information originating from multiple clinical modalities into coherent, context-aware recommendations. This capability has positioned LLMs as potential integrative intelligence platforms capable of bridging cardiovascular imaging, physiological signals, laboratory findings, and longitudinal patient records.^{12,34}

The emergence of multimodal foundation models further strengthens this concept. Contemporary cardiovascular AI is increasingly evolving toward systems capable of simultaneously processing echocardiographic images, ECG waveforms, biomarker profiles, radiological examinations, wearable sensor data, and narrative clinical documentation. Rather than replacing cardiologists, these systems aim to augment clinical reasoning by organizing complex information, identifying clinically relevant associations, and providing evidence-based recommendations consistent with contemporary clinical guidelines.^{2,4,10}

This progression from rule-based algorithms to multimodal LLMs represents more than a technological improvement; it signifies a conceptual transformation

in cardiovascular artificial intelligence. Whereas earlier AI systems functioned primarily as diagnostic tools for individual tasks, next-generation LLM-based platforms are increasingly being viewed as collaborative clinical reasoning systems capable of integrating diverse cardiovascular data streams into unified decision-support frameworks. Such an evolution constitutes an essential step toward precision cardiovascular medicine, where individualized diagnosis and treatment are informed by comprehensive multimodal information rather than isolated clinical variables.^{2,4,10}

4. Large language models in clinical medicine

Large language models (LLMs) represent one of the most significant advances in AI since the introduction of deep learning. Built upon transformer-based neural network architectures and trained on massive corpora of textual data, LLMs are capable of understanding, generating, summarizing, and reasoning over natural language with remarkable contextual awareness. Unlike traditional machine learning algorithms that require task-specific training, LLMs acquire broad linguistic and semantic knowledge during pretraining and can subsequently be adapted to diverse clinical applications through prompt engineering, fine-tuning, or retrieval-augmented generation (RAG) techniques.^{14,15}

The introduction of foundation models has fundamentally changed the role of AI in medicine. Previous AI systems were generally designed to solve narrowly defined tasks, such as detecting left ventricular dysfunction from echocardiographic images or identifying atrial fibrillation from electrocardiograms. Although these models often achieved expert-level performance, they lacked the ability to integrate heterogeneous clinical information or explain their conclusions within the broader clinical context. In contrast, LLMs are capable of processing structured and unstructured data simultaneously, allowing them to synthesize laboratory findings, imaging reports, clinical documentation, guideline recommendations, and patient history into coherent clinical narratives.^{13,14}

One of the most important characteristics of LLMs is their ability to interact with clinicians through natural language. Physicians no longer need to communicate with AI systems using predefined commands or specialized programming interfaces. Instead, complex clinical questions can be formulated conversationally, allowing the model to retrieve relevant knowledge, summarize evidence, generate differential diagnoses, explain clinical reasoning, and produce structured documentation. This natural language interface substantially lowers the barrier to AI adoption in routine clinical practice and facilitates integration into existing clinical workflows.^{14,16}

Representative contemporary LLM and multimodal model families relevant to cardiovascular medicine are summarized in Table 1. Because commercial systems evolve rapidly, the table reports the model family/version, developer, architecture and supported modalities, publicly disclosed parameter information, documented maximum context window, representative cardiovascular applications or evidence, principal advantages, and current limitations. Technical specifications are version-specific and were assessed through August 2026; where developers have not publicly disclosed parameter counts, this is stated explicitly.

Clinical documentation has emerged as one of the earliest and most practical applications of LLMs. Documentation-related activities account for a considerable proportion of physicians' daily workload and are recognized as a major contributor to professional burnout. LLMs have demonstrated the ability to generate discharge summaries, outpatient clinic letters, procedural reports, operative notes, and patient education materials while maintaining clinically appropriate language and standardized formatting. When combined with human review, these systems have the potential to improve documentation efficiency, reduce administrative burden, and increase physician time available for direct patient care.^{16,17}

Beyond documentation, LLMs are increasingly being investigated as clinical decision-support systems. Their ability to retrieve, summarize, and contextualize large volumes of medical literature enables rapid interpretation of evidence-based recommendations and contemporary clinical guidelines. Rather than replacing physician judgment, LLMs can assist clinicians by identifying relevant diagnostic pathways, highlighting potential therapeutic options, explaining guideline recommendations, and presenting individualized management strategies based on the available clinical information. Such capabilities are particularly valuable in complex clinical scenarios requiring simultaneous consideration of multiple comorbidities, medications, and diagnostic findings.^{14,15}

Medical education and scientific research constitute another rapidly expanding field of LLM implementation. Several studies have demonstrated competitive performance of advanced language models on medical licensing examinations, specialty board-style questions, and clinical reasoning assessments. In addition, LLMs are increasingly used for literature screening, manuscript editing, protocol development, hypothesis generation, and statistical code assistance. Nevertheless, these applications require continuous expert supervision because LLMs may generate inaccurate references, misinterpret scientific evidence, or produce plausible but factually incorrect statements.^{14,16,18}

Despite their remarkable capabilities, important limitations currently restrict the independent clinical use of LLMs. Hallucinations, defined as the generation of factually incorrect yet linguistically convincing information, remain among the most significant concerns. Furthermore, model performance depends heavily on input quality, prompt design, underlying training data, and access to updated medical knowledge. Additional challenges include algorithmic bias, insufficient explainability, limited transparency of proprietary models, patient privacy, cybersecurity, and regulatory compliance. Consequently, current consensus recommends that LLM-generated outputs should always undergo verification by qualified healthcare professionals before influencing clinical decision-making.^{16,17,19}

An emerging strategy to overcome several of these limitations involves combining LLMs with external knowledge retrieval systems, structured databases, and multimodal foundation models. Retrieval-augmented generation (RAG) enables language models to access current clinical guidelines, peer-reviewed literature, and institutional knowledge bases during inference, thereby reducing hallucinations and improving factual accuracy. Simultaneously, multimodal architectures extend language models beyond text by incorporating medical images, physiological signals, laboratory measurements, and longitudinal EHR data into a unified analytical framework. These advances suggest that the future role of LLMs will extend beyond conversational interfaces toward comprehensive clinical reasoning systems capable of integrating diverse sources of medical information.¹⁵

In this context, cardiovascular medicine represents an ideal environment for clinical implementation of LLMs. Cardiovascular decision-making routinely requires simultaneous interpretation of multimodal information derived from echocardiography, electrocardiography, biomarkers, imaging studies, and longitudinal clinical documentation. Rather than functioning as isolated chatbots, LLMs may serve as integrative intelligence platforms capable of organizing heterogeneous cardiovascular data into clinically meaningful recommendations. This conceptual transition—from language generation to multimodal clinical reasoning—forms the foundation for the next generation of precision cardiovascular medicine and represents the central framework discussed throughout this review.

5. Multimodal cardiovascular data

Modern cardiovascular medicine relies on the interpretation of multiple complementary sources of diagnostic information rather than any single examination.

Table 1. Contemporary LLM and multimodal model families relevant to cardiovascular medicine

Model / version	Developer	Architecture / modalities	Parameters	Maximum context size	Representative cardiovascular applications / evidence	Advantages	Current limitations
GPT-4 / GPT-4o	OpenAI	Transformer-based proprietary foundation models; GPT-4o supports multimodal text/image/audio workflows	Not publicly disclosed	Up to 128k tokens (version-dependent)	Clinical reasoning and decision support; documentation; guideline interpretation; multimodal image-based tasks; literature synthesis	Strong general reasoning; broad tool and multimodal integration; mature clinical-language capability	Hallucinations; proprietary architecture; prompt/version dependence; requires human verification; cardiovascular outcome validation remains limited
Gemini 2.5	Google DeepMind	Proprietary multimodal foundation model supporting text, images and other encoded modalities	Not publicly disclosed	Up to 1M tokens (model/version-dependent)	Multimodal clinical reasoning; ECG/image interpretation research; EHR and document synthesis	Very long context; native multimodal capability; strong cross-modal reasoning	Cardiovascular-specific prospective validation remains limited; proprietary system; performance is task- and version-dependent
Claude 4	Anthropic	Proprietary transformer-based language/multimodal model with long-context document reasoning	Not publicly disclosed	Up to 200k tokens (deployment-dependent)	Guideline and document synthesis; clinical documentation; multidisciplinary case summarization; reasoning support	Long-context analysis; strong document handling; useful for evidence synthesis	Limited cardiovascular-specific validation; proprietary architecture; not an autonomous clinical decision system
Llama 3	Meta	Open-weight decoder-only transformer family; primarily text-based, with local fine-tuning and deployment possible	8B and 70B in the original Llama 3 family	Up to 8k tokens for original Llama 3 releases	Institution-specific decision support; private EHR workflows; cardiovascular research and fine-tuned domain applications	Open weights; local deployment; institutional control and fine-tuning	Performance depends on model size and fine-tuning; original family is not natively multimodal; variable clinical accuracy
Med-PaLM 2	Google Research	Medical-domain adapted large language model based on the PaLM family; text-centered clinical knowledge and reasoning	Not publicly disclosed	Not consistently publicly specified for the research model	Medical question answering; evidence synthesis; clinical knowledge tasks relevant to cardiovascular decision support	Medical-domain adaptation; strong medical knowledge benchmarking	Limited public availability; text-centered rather than a universal raw-signal model; prospective cardiovascular validation required
Specialized cardiovascular LLMs / multimodal models*	Academic and institutional research	Heterogeneous domain-specific language, vision-language, waveform, or multimodal architectures	Varies by model	Varies by model	ECG interpretation, echocardiography reporting, PCI decision support, heart failure, inherited cardiomyopathy, and multimodal cardiovascular fusion	Task-specific adaptation; potential integration with validated cardiovascular pipelines	Heterogeneous methods and datasets; limited multicenter/prospective validation; limited generalizability; model-specific input constraints

Abbreviations: ECG: Electrocardiography; EHR, Electronic health record; LLM: Large language model. *Specialized cardiovascular LLMs include institution-specific and research-oriented models developed for cardiovascular applications and are not yet widely available for routine clinical practice.

Note: Parameter counts and context windows are reported only when publicly documented and are model/version-dependent. For proprietary systems, undisclosed technical details are identified as such. Specifications were assessed through August 2026.

Clinical decision-making requires simultaneous evaluation of imaging findings, physiological signals, laboratory biomarkers, and longitudinal patient records, each providing unique but incomplete insights into cardiovascular health. The increasing availability of digital healthcare data has accelerated the development of multimodal AI, which seeks to integrate heterogeneous information into a unified clinical framework. LLMs represent a promising advancement in this field because they are capable of contextualizing structured and unstructured data while supporting clinician-centered decision-making.^{1,9,20}

5.1. Echocardiography

Echocardiography remains the cornerstone of cardiovascular imaging owing to its wide availability, safety, portability, and ability to provide comprehensive anatomical and functional assessment in real time. It plays a central role in the diagnosis and management of heart failure, valvular heart disease, cardiomyopathies, congenital heart disease, pulmonary hypertension, infective endocarditis, and acute coronary syndromes. Standard transthoracic and transesophageal echocardiography provide extensive information regarding ventricular function, chamber dimensions, myocardial deformation, valve morphology, intracardiac pressures, and hemodynamic status.^{21,22}

Artificial intelligence (AI) has substantially enhanced echocardiographic analysis through automated image acquisition guidance, chamber segmentation, left ventricular ejection fraction estimation, strain analysis, valve assessment, and disease classification. Nevertheless, echocardiographic findings rarely provide sufficient information in isolation. Interpretation frequently depends on clinical history, ECG abnormalities, laboratory biomarkers, previous imaging examinations, and therapeutic response. Consequently, LLMs may serve as integrative platforms capable of combining quantitative echocardiographic measurements with broader clinical information to improve diagnostic interpretation and individualized patient management.^{9,22-24}

5.2. Electrocardiography

Electrocardiography is one of the most frequently performed diagnostic investigations in medicine and remains indispensable for the diagnosis of arrhythmias, myocardial ischemia, conduction abnormalities, electrolyte disturbances, inherited channelopathies, and numerous structural cardiac disorders. In addition to conventional diagnostic interpretation, advances in deep learning have demonstrated that ECG signals contain latent information capable of predicting left ventricular dysfunction, atrial fibrillation, hypertrophic cardiomyopathy, cardiac

amyloidosis, biological age, and future cardiovascular events.^{25,26}

Although AI-based ECG algorithms have achieved remarkable diagnostic accuracy, their outputs often require interpretation within the context of imaging findings, laboratory results, medications, and previous clinical history. LLMs offer a potential solution by synthesizing ECG-derived predictions with complementary clinical information, thereby facilitating more comprehensive cardiovascular assessment and supporting evidence-based clinical reasoning.^{1,26-28}

5.3. Cardiac CT and MRI

Cardiac CT and CMR have become essential components of contemporary cardiovascular diagnostics. Cardiac CT provides high-resolution visualization of coronary artery anatomy, plaque characteristics, coronary calcium burden, structural heart disease, and pre-procedural planning for structural interventions. In contrast, CMR offers unparalleled characterization of myocardial tissue, ventricular function, myocardial fibrosis, edema, inflammation, perfusion, and viability, making it the reference standard for numerous cardiomyopathies and inflammatory heart diseases.^{29,30}

The increasing complexity of advanced cardiovascular imaging generates large volumes of quantitative and qualitative information that may be challenging to integrate during routine clinical practice. LLMs have the potential to summarize radiological reports, correlate imaging findings with echocardiographic and ECG abnormalities, incorporate biomarker profiles, and generate clinically meaningful interpretations that support multidisciplinary decision-making while preserving physician oversight.^{13,29}

5.4. Laboratory biomarkers

Laboratory biomarkers constitute another fundamental component of cardiovascular diagnosis, prognosis, and therapeutic monitoring. Cardiac troponins remain the gold standard biomarkers for myocardial injury, whereas natriuretic peptides are central to the diagnosis and management of heart failure. Additional biomarkers—including C-reactive protein, D-dimer, lipoprotein(a), high-sensitivity C-reactive protein, galectin-3, soluble ST2, and various inflammatory mediators—provide complementary information regarding inflammation, fibrosis, thrombosis, and cardiovascular risk stratification.^{31,32}

Because biomarker interpretation depends on timing, serial measurements, renal function, age, comorbidities, and accompanying imaging findings, isolated laboratory values rarely provide complete clinical information. By integrating biomarker trajectories with imaging, ECG

findings, medications, and longitudinal EHR data, LLMs may facilitate more individualized diagnostic assessment and improve interpretation of complex cardiovascular presentations.^{1,31,33}

5.5. Electronic health records

Electronic health records (EHRs) represent the central repository of longitudinal clinical information and contain a combination of structured data, including diagnoses, medications, laboratory values, and vital signs, together with extensive unstructured documentation such as physician notes, discharge summaries, operative reports, and consultation letters. Although EHRs contain valuable clinical knowledge, much of this information remains difficult to access efficiently because of fragmented documentation and information overload.^{5,6}

Natural language processing has substantially improved extraction of clinically relevant information from EHRs, while recent LLMs have demonstrated superior capabilities in summarizing longitudinal patient histories, identifying clinically significant events, supporting documentation, and retrieving guideline-relevant information. Importantly, EHRs serve as the central framework linking all other cardiovascular data sources, including echocardiography, ECG, laboratory biomarkers, CT, and CMR. Therefore, effective integration of EHR-derived information with multimodal diagnostic data represents a critical prerequisite for precision cardiovascular medicine and constitutes one of the most promising applications of LLM-based clinical intelligence.^{1,5,6}

5.6. Biomedical sensors, data acquisition, and signal quality

Multimodal cardiovascular intelligence begins with data acquisition rather than with the language model itself. Physiological information derived from ECG electrodes, wearable photoplethysmography, blood pressure sensors, inertial sensors, implantable devices, and other biomedical interfaces must be acquired with adequate sampling, synchronization, calibration, signal-to-noise ratio, and metadata integrity before downstream AI can operate reliably. Motion artifact, electrode-skin impedance, missing channels, clock drift, device-specific preprocessing, and incomplete temporal alignment can propagate through the analytical pipeline and produce apparently coherent but clinically misleading downstream outputs. Accordingly, sensor quality control, signal preprocessing, temporal synchronization, and provenance should be regarded as explicit components of an integrative intelligence architecture rather than as invisible preprocessing steps.

Experimental multimodal systems illustrate this

principle. Bibbò *et al.*⁶⁸ combined synchronized EEG and EMG acquisition with deep fusion in a wearable human-machine interface and reported real-time classification performance, demonstrating how acquisition quality, synchronization, and modality-specific preprocessing precede intelligent fusion. Although this system was developed for an exoskeleton rather than cardiovascular care, it provides a useful engineering example of the acquisition-to-fusion pipeline highlighted by the reviewer. In cardiovascular applications, the same principle applies: an LLM or orchestration layer cannot compensate reliably for corrupted, poorly synchronized, or clinically unrepresentative source signals.⁶⁸

5.7. Multimodal fusion strategies

Multimodal fusion can occur at several computational levels. Early or data-level fusion combines aligned inputs or minimally processed representations before substantial modality-specific feature extraction; it can exploit low-level interactions but is particularly sensitive to scale differences, missing data, noise, and imperfect temporal or spatial alignment. Intermediate or feature-level fusion first encodes each modality separately and then combines learned representations in a shared latent space. This strategy permits richer cross-modal interactions and can use concatenation, gating, joint embeddings, co-attention, or cross-attention mechanisms. Late or decision-level fusion preserves independent modality-specific models and combines their predictions, scores, or structured outputs. Late fusion is often easier to audit and can retain validated specialist models, but it may capture fewer low-level interactions between modalities.

Attention-based architectures increasingly support intermediate fusion by allowing features from one modality to selectively condition representations from another. Cross-attention is particularly relevant when modalities differ substantially in dimensionality or temporal structure. In a recent cardiovascular example, MAF-Net combined clinical variables with ECG-derived features through bidirectional cross-attention and reported improved classification performance on the PTB-XL dataset compared with unimodal or alternative fusion approaches. Such results support the value of heterogeneous data fusion, but they should not be extrapolated directly to end-to-end LLM-based clinical decision support because the task, architecture, and validation setting are different.⁶⁹

For the integrative intelligence framework proposed here, no single fusion strategy should be assumed to be universally optimal. Raw waveforms and images may be processed by dedicated encoders or validated specialist algorithms; intermediate representations may be fused by

multimodal foundation models; and higher-level outputs may be combined through an LLM-based orchestration layer with EHR context, biomarkers, retrieved evidence, and physician input. The appropriate fusion level should therefore depend on the clinical task, data quality, temporal alignment, availability of validated modality-specific models, and the need for traceability and failure isolation. This layered interpretation also prevents conflation of multimodal signal fusion with language model reasoning: the former concerns how heterogeneous data are computationally combined, whereas the latter concerns how validated representations and evidence are synthesized into a clinically interpretable output.

Collectively, these cardiovascular data domains, together with the acquisition and fusion layers described above, illustrate the inherently multimodal nature of cardiovascular medicine. Each modality contributes complementary information, yet none is independently sufficient for optimal clinical decision-making. The ability of LLMs to synthesize imaging findings, physiological signals, laboratory biomarkers, and longitudinal clinical documentation into coherent, patient-specific recommendations represents a major conceptual advance over conventional task-specific AI systems. This multimodal integration forms the basis of the proposed framework of LLMs as integrative intelligence, discussed in the following section.

6. LLMs as integrative intelligence: Definition and conceptual novelty

The term “integrative intelligence” is used in this review as a clinical systems concept rather than as the name of a new neural-network architecture. It describes an orchestration layer in which heterogeneous cardiovascular information is assembled from multiple sources, the outputs of modality-specific analytical systems are contextualized with clinical text, laboratory data, longitudinal records, and external evidence, and the resulting synthesis is presented to the clinician for verification and decision-making. The proposed framework therefore does not assume that a conventional text-only LLM can directly interpret every raw cardiovascular modality.

A strict distinction is required among four related but non-equivalent technologies. First, text-based LLMs primarily process natural-language inputs and are well suited to tasks such as information extraction, summarization, documentation, guideline retrieval, and synthesis of textual clinical evidence. Second, large multimodal or vision language models extend foundation-model capabilities to combinations of text with images or other encoded modalities, but their supported inputs

and performance depend on the specific architecture, training data, and validation setting. Third, dedicated cardiovascular AI algorithms remain necessary for many modality-specific tasks, including ECG waveform analysis, echocardiographic segmentation and quantification, CT or CMR image analysis, and physiological-signal processing. Fourth, an LLM-based orchestration layer can receive validated outputs from these specialized systems; combine them with EHR data, biomarkers, reports, and retrieved evidence; and organize the information into a clinically interpretable decision-support representation.

Accordingly, references in this review to multimodal cardiovascular integration should not be interpreted as implying that a standard text-only LLM directly analyzes raw echocardiographic cine loops, ECG waveforms, CT datasets, CMR sequences, or continuous physiological signals. Direct analysis of such data requires an appropriately trained multimodal foundation model or a dedicated modality-specific algorithm. Within the integrative intelligence framework, the LLM may instead function as the coordinating and reasoning interface that synthesizes the outputs of those analytical components with patient context and current evidence.

Conceptually, integrative intelligence differs from multimodal foundation models, conventional multimodal decision-support systems, and generic AI orchestration. A multimodal foundation model is principally a model architecture trained to learn transferable representations across more than one data type. A multimodal clinical decision-support system may combine several data streams for a predefined prediction or classification task. AI orchestration refers more broadly to coordinating models, tools, retrieval systems, or agents. Integrative intelligence, as proposed here, is narrower in clinical purpose but broader in informational scope: it places orchestration within a physician-supervised cardiovascular workflow and explicitly links modality-specific analysis, longitudinal patient context, evidence provenance, uncertainty, and human clinical judgment. Its proposed novelty therefore lies in the organization of these components into a unified clinical reasoning framework, not in claiming a novel foundation-model architecture.

This distinction also defines the evidentiary status of the framework. Individual components—including dedicated ECG and imaging algorithms, multimodal cardiovascular foundation models, LLM-based extraction from cardiovascular reports, and retrieval-supported clinical language systems—have been evaluated to varying degrees. However, the complete integrative intelligence framework described here has not yet been prospectively validated as a single end-to-end cardiovascular intervention. It

should therefore be regarded as a translational framework that organizes existing and emerging technologies and generates testable hypotheses for prospective clinical evaluation, rather than as an already proven improvement in cardiovascular outcomes.

The proposed concept of LLMs as integrative intelligence is illustrated in Figure 1, which demonstrates how heterogeneous cardiovascular data are transformed into integrated clinical intelligence supporting physician decision-making.

6.1. Data fusion and orchestration

One of the greatest challenges in cardiovascular medicine is the fragmentation of clinical information. Patient management routinely depends on data originating from multiple independent systems, including echocardiography, electrocardiography, cardiac CT, CMR, laboratory information systems, wearable devices, and EHRs. Although each modality contributes valuable diagnostic information, clinicians remain responsible for integrating these heterogeneous datasets into a unified clinical interpretation.^{35,36,37}

LLMs provide a unique opportunity to overcome this fragmentation through multimodal data fusion. Rather than analyzing individual variables independently, LLMs can contextualize quantitative measurements, narrative reports, longitudinal trends, and physician documentation within a single clinical framework. For example, reduced left ventricular ejection fraction identified on echocardiography may be interpreted together with elevated natriuretic peptides, chronic kidney disease

documented in the EHR, atrial fibrillation detected on ECG, and previous coronary angiography findings to generate a comprehensive clinical summary that reflects the patient's overall cardiovascular status rather than isolated diagnostic observations.^{13,34,38}

6.2. Clinical reasoning under physician oversight

Clinical reasoning remains one of the most complex cognitive processes in medicine. Experienced cardiologists continuously integrate symptoms, medical history, physical examination, diagnostic testing, therapeutic response, and guideline recommendations before establishing a diagnosis or selecting treatment. Conventional AI algorithms generally address only one component of this process, whereas LLMs offer the possibility of supporting sequential clinical reasoning across multiple stages of patient evaluation.

Rather than functioning as autonomous diagnosticians, LLMs may assist physicians by organizing available evidence, generating differential diagnoses, identifying missing diagnostic information, highlighting inconsistencies between clinical findings, and explaining the rationale behind potential management strategies. Such collaborative reasoning is particularly valuable in patients with multiple comorbidities, atypical presentations, or conflicting diagnostic findings, where cognitive integration rather than isolated prediction determines clinical success. Nevertheless, current evidence consistently indicates that LLM-generated reasoning requires expert verification because model performance remains sensitive to incomplete data, prompt design, and factual hallucinations.^{16,34,39}

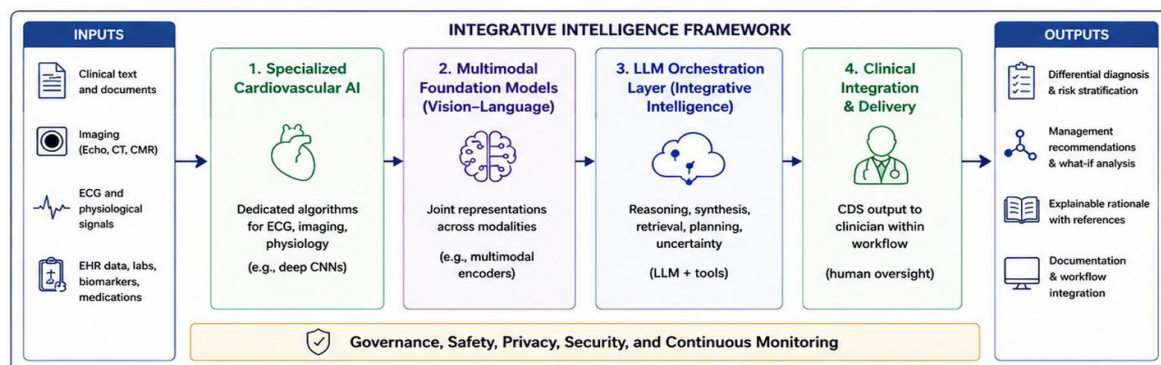


Figure 1. Integrative intelligence architecture for cardiovascular medicine. Heterogeneous cardiovascular data are first processed by appropriate modality-specific or multimodal AI components. An LLM-enabled orchestration layer subsequently integrates validated outputs with clinical context and evidence to support physician-supervised decision-making. The framework does not imply that a conventional text-based LLM directly interprets raw ECG waveforms or cardiovascular images. Schematic created by the author.

Abbreviations: AI: Artificial intelligence; CDS: Clinical decision support; CMR: Cardiovascular magnetic resonance; CNN: Convolutional neural network; CT: Computed tomography; ECG: Electrocardiography; LLM: Large language model.

6.3. Guideline interpretation

Cardiovascular medicine is characterized by rapidly evolving evidence-based clinical practice guidelines issued by professional societies such as the European Society of Cardiology (ESC) and the American College of Cardiology/American Heart Association (ACC/AHA). The increasing complexity and volume of these recommendations make their consistent implementation challenging in everyday clinical practice.

LLMs may facilitate guideline implementation by rapidly retrieving relevant recommendations, summarizing evidence supporting specific therapeutic decisions, identifying appropriate diagnostic pathways, and adapting recommendations to individual clinical scenarios. Integration with retrieval-augmented generation (RAG) systems allows language models to reference updated guideline documents rather than relying exclusively on pretrained knowledge, thereby improving factual accuracy while reducing hallucinations. Importantly, LLMs should be viewed as decision-support systems that assist clinicians in navigating contemporary evidence rather than replacing independent clinical judgment.^{6,16,39}

6.4. Report generation

Clinical documentation represents one of the most time-consuming aspects of cardiovascular practice. Echocardiographic reports, catheterization summaries, outpatient clinic letters, discharge summaries, multidisciplinary meeting reports, and patient education materials require substantial physician time while often containing repetitive information.

LLMs have demonstrated remarkable ability to generate structured medical reports using standardized terminology and clinically appropriate language. When integrated with hospital information systems, they may automatically summarize imaging findings, laboratory trends, procedural results, and physician documentation into coherent reports requiring only clinician verification before final approval. Such applications have the potential to reduce administrative burden, improve documentation consistency, and increase time available for direct patient care while maintaining physician responsibility for clinical accuracy.^{16,17,39}

6.5. Patient-specific decision support

Perhaps the greatest translational potential of LLMs lies in personalized clinical decision support. Precision cardiovascular medicine increasingly requires individualized therapeutic recommendations that account for patient characteristics, multimorbidity, longitudinal disease progression, medication history, imaging findings,

biomarkers, lifestyle factors, and patient preferences.

Within the proposed integrative intelligence framework, LLMs function as orchestrators that combine these heterogeneous inputs into patient-specific recommendations presented transparently to the treating physician. Rather than generating a single diagnostic answer, the model may provide ranked diagnostic possibilities, summarize supporting evidence, identify missing information, estimate uncertainty, and explain how current guideline recommendations apply to the individual patient. This collaborative approach preserves physician autonomy while enhancing clinical efficiency, consistency, and evidence-based practice. Importantly, shared decision-making remains central to cardiovascular care, and LLM-generated recommendations should always be interpreted within the context of patient values, preferences, and clinician expertise.^{13,34,40}

Collectively, these capabilities redefine the role of large language models in cardiovascular medicine. Instead of functioning merely as advanced conversational agents, LLMs can be conceptualized as integrative intelligence systems that bridge multimodal diagnostics, clinical reasoning, evidence synthesis, and patient-centered decision support. This conceptual framework extends beyond traditional task-specific artificial intelligence and provides a translational model for future multimodal cardiovascular care in which human expertise and AI operate collaboratively rather than competitively.

7. Clinical applications

The clinical value of LLMs ultimately depends on their ability to improve real-world patient care. Cardiovascular medicine provides an ideal environment for LLM implementation because diagnosis and treatment routinely require integration of multimodal information originating from imaging, physiological signals, laboratory biomarkers, clinical documentation, and evidence-based guidelines. Rather than replacing existing AI algorithms, LLMs may serve as an integrative clinical intelligence layer that supports physicians throughout the diagnostic and therapeutic pathway. Several cardiovascular conditions illustrate the translational potential of this approach. Importantly, most proposed cardiovascular applications remain investigational; potential clinical benefits should therefore be interpreted as hypotheses for prospective validation rather than as established improvements in patient outcomes.^{7,41}

The proposed clinical workflow is illustrated in [Figure 2](#), demonstrating how LLMs integrate multimodal cardiovascular data throughout the patient journey, from initial clinical assessment to individualized treatment decisions and longitudinal follow-up.

7.1. Acute coronary syndrome

Acute coronary syndrome is a time-critical cardiovascular emergency requiring rapid integration of clinical presentation, electrocardiographic findings, cardiac biomarkers, imaging studies, coronary anatomy, and current guideline recommendations. Despite well-established diagnostic algorithms, considerable heterogeneity exists among patients presenting with chest pain, particularly in cases of non-ST-elevation myocardial infarction, myocardial infarction with non-obstructive coronary arteries (MINOCA), or atypical clinical presentations.^{29,31}

Large language models (LLMs) may enhance acute coronary syndrome management by simultaneously interpreting emergency department documentation, serial high-sensitivity cardiac troponin measurements, ECG reports, echocardiographic findings, coronary CT angiography, previous cardiovascular history, medications, and risk scores. Integrated with retrieval-augmented generation (RAG), LLMs can rapidly summarize current European Society of Cardiology (ESC) and American College of Cardiology/American Heart Association (ACC/AHA) recommendations, identify appropriate diagnostic pathways, and support individualized treatment planning while maintaining physician oversight. Such systems may also facilitate communication between emergency physicians, cardiologists, and catheterization laboratories through automated generation of structured clinical

summaries and transfer documentation.^{29,34,42}

7.2. Heart failure

Heart failure represents one of the most complex chronic cardiovascular syndromes because diagnosis and management depend on continuous integration of multimodal information. Clinical assessment incorporates symptoms, physical examination, natriuretic peptides, echocardiographic parameters, ECG findings, renal function, comorbidities, medication history, and repeated longitudinal follow-up. Consequently, heart failure management is particularly well-suited for LLM-supported clinical reasoning.^{38,43}

LLMs may assist clinicians by integrating quantitative echocardiographic measurements—including left ventricular ejection fraction, global longitudinal strain, diastolic function, right ventricular performance, and valvular abnormalities—with laboratory biomarkers, medication optimization, device therapy eligibility, and contemporary guideline recommendations. Furthermore, longitudinal analysis of serial EHR data may enable early identification of clinical deterioration, medication non-adherence, repeated hospitalizations, or disease progression. Such comprehensive patient-specific assessment may support personalized management strategies while reducing documentation burden and improving multidisciplinary communication.^{38,43,44}

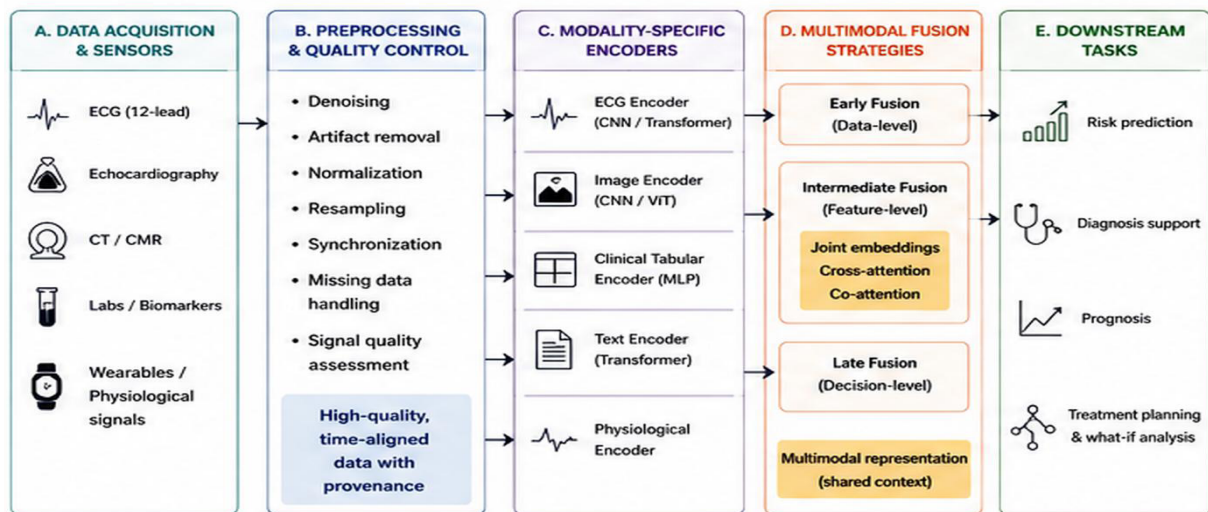


Figure 2. Multimodal data acquisition, processing, and fusion pipeline. The diagram presents a conceptual pipeline illustrating the progression from biomedical data acquisition and quality control through modality-specific processing and multimodal fusion to clinical integration. Fusion may occur at the data, feature, or decision level, depending on the clinical task, modality characteristics, and validation requirements. The figure emphasizes that reliable downstream artificial intelligence depends on appropriate acquisition, preprocessing, synchronization, and data quality. Schematic created by the author. Abbreviations: CMR: Cardiovascular magnetic resonance; CNN: Convolutional neural network; CT: Computed tomography; ECG: Electrocardiography; MLP: Multilayer perceptron; ViT: Vision transformer.

7.3. Arrhythmias

Artificial intelligence has already demonstrated excellent performance in automated ECG interpretation and arrhythmia detection. However, clinical management of rhythm disorders extends beyond rhythm classification alone. Appropriate therapeutic decisions require consideration of structural heart disease, ventricular function, ischemic burden, inherited disorders, anticoagulation status, imaging findings, and individual patient characteristics.^{25,45,46}

Within the proposed integrative intelligence framework, LLMs may combine outputs generated by AI-enabled ECG algorithms with echocardiographic findings, cardiac MRI reports, laboratory biomarkers, electrophysiology studies, and longitudinal EHR information. Such integration may facilitate differential diagnosis of complex arrhythmias, identify reversible causes, evaluate stroke and bleeding risk, summarize indications for catheter ablation or implantable cardiac devices, and generate individualized management recommendations based on current international guidelines. This collaborative model has particular relevance for atrial fibrillation, ventricular arrhythmias, inherited channelopathies, and cardiac implantable electronic device follow-up.^{34,45}

7.4. Valvular heart disease

The diagnosis and treatment of valvular heart disease increasingly depend on multimodal imaging and multidisciplinary Heart Team decision-making. Echocardiography remains the primary imaging modality, while cardiac CT, cardiac MRI, coronary angiography, surgical risk assessment, frailty evaluation, and patient preferences all contribute to therapeutic planning.^{47,48}

LLMs may support Heart Team discussions by integrating imaging reports, procedural planning data, operative risk scores, laboratory findings, and guideline recommendations into concise patient-specific summaries. Automated synthesis of serial echocardiographic examinations may facilitate identification of disease progression and optimal timing of intervention, while integration of CT-derived anatomical measurements may improve procedural planning for transcatheter valve interventions. By organizing complex clinical information into structured decision-support reports, LLMs have the potential to improve consistency and efficiency of multidisciplinary cardiovascular care.^{13,47}

7.5. Cardio-oncology

Cardio-oncology represents one of the fastest-growing subspecialties in cardiovascular medicine and exemplifies the necessity for multimodal clinical integration. Patients

receiving potentially cardiotoxic cancer therapies require continuous assessment of cardiovascular risk factors, cardiac biomarkers, echocardiographic surveillance, ECG monitoring, oncological treatment history, and long-term follow-up.^{49,50}

Large language models (LLMs) may facilitate collaboration between cardiologists and oncologists by integrating imaging findings, serial biomarker measurements, chemotherapy protocols, radiotherapy exposure, and cardiovascular risk profiles into unified clinical summaries. Such systems may assist in identifying patients at increased risk of cancer therapy-related cardiac dysfunction, recommending surveillance intervals according to international cardio-oncology guidelines, generating standardized reports, and supporting individualized preventive strategies. Because management decisions frequently require simultaneous consideration of oncological efficacy and cardiovascular safety, LLM-assisted multimodal reasoning may substantially enhance multidisciplinary decision-making in this rapidly evolving field.^{49,50}

To distinguish established evidence from proof-of-concept and future applications, representative quantitative cardiovascular studies are summarized in [Table 2](#). The available literature spans randomized clinician-assistance trials, retrospective model comparisons, report-generation studies, multimodal fusion experiments, and systematic reviews; consequently, reported performance metrics should be interpreted within the design and validation limits of each study.^{4,7,8,24,27,46,69}

The principal clinical applications of LLMs across the cardiovascular care continuum are summarized in [Table 3](#).

Collectively, these clinical applications illustrate that the greatest value of LLMs does not arise from isolated diagnostic predictions but from their ability to integrate heterogeneous cardiovascular information into clinically meaningful, patient-specific recommendations. Acute coronary syndrome, heart failure, arrhythmias, valvular heart disease, and cardio-oncology all require continuous synthesis of multimodal diagnostic data and evidence-based recommendations. Within this context, LLMs may function as collaborative clinical intelligence systems that enhance physician efficiency, improve information accessibility, and facilitate precision cardiovascular medicine while preserving the central role of clinician expertise in final decision-making.

8. Human–AI collaborative decision making

The rapid evolution of LLMs has generated considerable discussion regarding the future relationship between AI and healthcare professionals. Although recent advances have

Table 2. Representative quantitative evidence for LLMs and multimodal AI relevant to cardiovascular decision support

Study / model	CV domain and task	Input modality	Sample and design	Comparator / reference standard	Key quantitative findings	External validation	Evidence status / main limitations
O'Sullivan <i>et al.</i> , 2026 ⁵ ; AMIE	Inherited cardiomyopathy; diagnostic and management support	AMIE received clinical text reports derived from multimodal cardiac testing; cardiologists also reviewed raw ECG, TTE, CMR, Holter, and CPX data	107 real-world cases; randomized, counterbalanced clinician-assistance trial; 9 general cardiologists; blinded subspecialist evaluation	General cardiologists with AMIE vs general cardiologists without AMIE	AMIE-assisted assessments preferred 46.7% vs 32.7% ($p=0.02$); clinically significant errors 13.1% vs 24.3% ($p=0.033$); missing content 17.8% vs 37.4% ($p=0.0021$); time saved in 50.5% of cases	No independent external cohort; real-world cases from a subspecialty practice	Strongest current clinical evidence in this table, but retrospective patient data, single subspecialty domain, clinician-in-the-loop design, and 6.5% clinically significant hallucination rate reported by users
Lee <i>et al.</i> , 2025 ²⁷ ; GPT-4o, Gemini 2.5 Pro, ECG Buddy	Myocardial infarction detection from ECG images	12-lead ECG images only	928 ECG recordings; retrospective comparative study; 239 MI-positive and 689 MI-negative	Dedicated ECG AI (ECG Buddy); dataset labels	GPT-4o: accuracy 65.95%, AUC 57.34%, sensitivity 36.40%, specificity 76.20%; Gemini 2.5 Pro: accuracy 29.63%, AUC 51.63%; ECG Buddy: accuracy 96.98%, AUC 98.8%, sensitivity 96.65%, specificity 97.10%	No	Demonstrates clear superiority of dedicated ECG AI over general-purpose multimodal LLMs for this task; single public image dataset, no ancillary clinical context, retrospective design
Chao <i>et al.</i> , 2025 ²⁴ ; EchoGPT	Echocardiography report summarization	Structured and free-text echocardiography findings / reports	97,506 development reports; 1,000-report Arizona validation set; 1,000-report MIMIC-EchoNotes external dataset; retrospective	Open-source LLM baselines and human expert summaries	Fine-tuned EchoGPT: ROUGE-L 55.7±17.8, BERTScore 91.6±3.0; expert review found greater conciseness than human summaries ($p<0.001$) with no significant difference in completeness, correctness, or clinical utility	Yes; MIMIC-EchoNotes, with performance drop (RadGraph F1 47.7 to 25.1)	Clinically relevant reporting task with external testing, but retrospective, report-style dependent, and developed largely in a predominantly White Mayo population
Lin <i>et al.</i> , 2026 ⁶ ; 15 contemporary LLMs	Coronary disease; PCI decision support	Structured clinical and coronary evaluation data across nine decision tasks	93 patients with moderate-to-severe coronary stenosis; retrospective single-center study	PCI indication established by consensus of 3 cardiologists	Llama-3.3-70B-Instruct sensitivity 99.6%; Doubao-1.5-Pro-32k specificity 78.9%; Claude-3.7-Sonnet accuracy 0.95 and F1 0.97 on task T4; substantial between-model and task-dependent variation	No	Proof-of-concept clinical decision-support evidence; small single-center cohort, retrospective design, model/version dependence, and marked sensitivity to task formulation and patient age

(Cont'd...)

Table 2. (Continued)

Study / model	CV domain and task	Input modality	Sample and design	Comparator / reference standard	Key quantitative findings	External validation	Evidence status / main limitations
Qu <i>et al.</i> , 2026 ⁶ ; MAF-Net	Multimodal cardiovascular disease classification	ECG signals plus structured clinical features	PTB-XL dataset; repeated experimental evaluation; five diagnostic superclasses	Unimodal X branch, ECG-only Y branch, and published baseline models	Multimodal model accuracy 90.75%±0.32%, F1 0.8069±0.005, AUC 0.9407±0.002; ECG-only branch accuracy 88.86%±0.34%	No	Validated multimodal fusion proof-of-concept, but not an LLM; single public dataset and no prospective or real-world external validation
Gendler <i>et al.</i> , 2026 ⁴ ; systematic review	Broad cardiology LLM applications	Text-based questions, educational content, diagnostic and acute-care tasks across included studies	33 studies with quantitative outcomes; systematic review of PubMed and Scopus literature	Varied expert, guideline, or dataset reference standards	Examples included 91% accuracy for ChatGPT-3.5 on 47 heart-failure questions; Bing Chat omitted key MI first-aid steps in 25%–45% of cases	Not applicable at review level	Best synthesis of the cardiology literature to date, but marked heterogeneity in tasks, prompts, model versions, outcomes, and generally limited prospective validation
Ferreira Santos <i>et al.</i> , 2025 ⁵ ; systematic review	CVD prevention, patient education, and clinical decision support	Predominantly text-based LLM interactions	35 observational studies; systematic review	Varied clinical guidelines and expert assessments	72% of studies addressed patient education, 17% clinical decision support, and 11% both; 91% evaluated ChatGPT	Not applicable at review level	Shows breadth of use but predominantly observational evidence, frequent high risk of bias, limited individualized-care validation, and sparse outcome data

Notes: Evidence status reflects the design and validation setting of the cited study and should not be interpreted as a formal grading system.

Abbreviations: AUC: Area under the receiver operating characteristic curve; CMR: Cardiovascular magnetic resonance; CPX: Cardiopulmonary exercise testing; CV: Cardiovascular; ECG: Electrocardiography; F1: F1 score; LLM: Large language model; MI: Myocardial infarction; PCI: Percutaneous coronary intervention; TTE: Transthoracic echocardiography.

Table 3. Clinical applications and potential value of LLM-enabled integrative intelligence in cardiovascular medicine

Clinical domain	Multimodal data sources	LLM functions	Expected clinical benefits
Acute coronary syndrome	ECG, hs-cTn, echocardiography, coronary CT, EHR	Differential diagnosis, guideline interpretation, report generation	Potential to support earlier diagnosis, faster reperfusion decisions, standardized care
Heart failure	Echocardiography, BNP/NT-proBNP, ECG, laboratory data, EHR	Phenotyping, therapy optimization, longitudinal monitoring	Potential to support personalized treatment, reduced readmissions, improved prognosis
Arrhythmias	ECG, Holter monitoring, cardiac MRI, EHR	Rhythm interpretation, risk stratification, ablation support	Potential to support improved arrhythmia management and anticoagulation decisions
Valvular heart disease	Echocardiography, CT, MRI, Heart Team reports	Disease severity assessment, procedural planning, documentation	Potential to support better timing of intervention and multidisciplinary communication
Cardio-oncology	Echocardiography, troponin, natriuretic peptides, oncology records	Surveillance planning, toxicity monitoring, risk prediction	Potential to support earlier detection of cardiotoxicity and individualized follow-up
Preventive cardiology	Risk factors, laboratory biomarkers, wearables, lifestyle data	Cardiovascular risk prediction, preventive recommendations	Potential to support earlier intervention and personalized prevention
Cardiac intensive care	Continuous ECG, hemodynamic monitoring, laboratory results, EHR	Real-time data summarization, clinical alerts, decision support	Potential to support improved workflow and faster recognition of clinical deterioration

Abbreviations: BNP: B-type natriuretic peptide; CT: computed tomography; ECG: electrocardiography; EHR: electronic health record; hs-cTn: High-sensitivity cardiac troponin; LLM: large language model; MRI: Magnetic resonance imaging; NT-proBNP: N-terminal pro-B-type natriuretic peptide. Notes: Potential or anticipated clinical value is presented as hypothesis-generating unless supported by prospective clinical outcome data. Terms such as earlier diagnosis, faster reperfusion, fewer readmissions, improved prognosis, or better survival should not be interpreted as established effects of LLM-supported care.

demonstrated impressive capabilities in clinical reasoning, documentation, evidence retrieval, and multimodal data integration, current evidence consistently indicates that LLMs should not be viewed as autonomous decision-makers. Instead, their greatest clinical value lies in augmenting physician expertise through collaborative decision support. We propose that the future of cardiovascular medicine will be defined not by physician replacement but by Human-AI Collaborative Decision Making, in which clinicians and LLMs contribute complementary strengths to optimize patient care.^{13,16,34,51,52,53}

8.1. Physician-in-the-loop

Cardiovascular medicine involves complex diagnostic and therapeutic decisions that frequently require interpretation of incomplete, uncertain, or conflicting information. While LLMs can rapidly synthesize multimodal data, physicians contribute contextual understanding, ethical judgment, clinical intuition, procedural expertise, and responsibility for patient care. Consequently, clinical implementation should follow a physician-in-the-loop model, in which LLM-generated analyses are reviewed, interpreted, and validated by qualified cardiologists before influencing patient management.^{16,39,54}

Rather than functioning independently, LLMs should

operate as intelligent clinical assistants that continuously support—but never replace—human expertise. This collaborative approach preserves physician accountability while enabling faster access to relevant clinical information, more consistent guideline implementation, and improved workflow efficiency.

8.2. Explainable clinical recommendations

Trust represents one of the most important prerequisites for successful implementation of AI in cardiovascular medicine. Physicians are unlikely to rely on recommendations that cannot be adequately explained or justified. Therefore, future LLM-based systems should prioritize transparency by clearly documenting the evidence supporting every recommendation.

Instead of providing isolated diagnostic conclusions, integrative LLM platforms should summarize the clinical findings contributing to each recommendation, identify the relevant guideline statements, acknowledge diagnostic uncertainty, and explicitly distinguish factual evidence from probabilistic inference. Explainable outputs facilitate physician verification, improve clinician confidence, and support safe implementation in routine cardiovascular practice.^{16,55,56}

8.3. Shared decision-making

Modern cardiovascular care increasingly emphasizes shared decision-making between physicians and patients. Therapeutic choices such as coronary revascularization, valve intervention, implantable cardioverter-defibrillator implantation, catheter ablation, or advanced heart failure therapies often require careful consideration of patient preferences, quality of life, procedural risks, and expected outcomes.

LLMs may contribute to this process by translating complex medical terminology into understandable language, generating individualized educational materials, summarizing benefits and risks of therapeutic alternatives, and facilitating communication between clinicians and patients. Nevertheless, treatment decisions should always remain collaborative processes involving informed patients and experienced healthcare professionals rather than algorithm-driven recommendations alone.^{40,57}

8.4. Clinical responsibility and ethical accountability

Despite remarkable advances in generative AI, professional responsibility for diagnosis and treatment remains exclusively with healthcare professionals. LLMs should therefore be regarded as clinical decision-support systems rather than autonomous medical practitioners.

Clear governance frameworks are required to define responsibilities regarding AI-generated recommendations, documentation, data quality, model updating, and error management. Institutions implementing LLM-supported clinical workflows should establish mechanisms for human verification, continuous quality monitoring, auditability, and reporting of model failures. Such governance will be essential for maintaining patient safety, legal accountability, and public trust as generative AI becomes increasingly integrated into healthcare systems.^{16,17,55}

8.5. Workflow integration and multidisciplinary collaboration

The successful implementation of LLMs depends not only on model performance but also on seamless integration into existing clinical workflows. Cardiovascular care routinely involves multidisciplinary collaboration among cardiologists, cardiac surgeons, imaging specialists, emergency physicians, intensivists, oncologists, nurses, pharmacists, and primary care physicians. Fragmented communication between these professionals contributes to diagnostic delays, duplicated documentation, and inconsistent information exchange.

Large language models (LLMs) have the potential to function as intelligent communication hubs that organize

multimodal clinical information into standardized summaries accessible to all members of the multidisciplinary team. Automated synthesis of imaging reports, laboratory trends, ECG findings, medications, procedural history, and longitudinal EHR documentation may improve continuity of care while reducing administrative burden. Such integration could be particularly valuable in Heart Team discussions, cardio-oncology conferences, heart failure programs, and emergency cardiovascular care, where timely access to comprehensive patient information is critical for optimal decision-making.^{13,34,39,58}

Ultimately, the concept of Human-AI Collaborative Decision Making recognizes that the strengths of AI and human clinicians are complementary rather than competitive. LLMs excel at processing large volumes of heterogeneous information, retrieving evidence, organizing clinical documentation, and supporting multimodal reasoning, whereas physicians provide contextual interpretation, ethical judgment, empathy, communication, and accountability. The future of cardiovascular medicine is therefore unlikely to involve autonomous AI replacing cardiologists; instead, it will be characterized by collaborative intelligence in which LLMs enhance physician capabilities while preserving human oversight as the central element of safe, patient-centered care.

9. Safety, reliability, validation, and governance

The translation of LLM-enabled cardiovascular decision support from experimental systems to routine care requires a safety framework that extends beyond average benchmark accuracy. High-stakes use demands attention to data quality, uncertainty, omission errors, prompt sensitivity, automation bias, model drift, subgroup performance, out-of-distribution behavior, retrieval failure, cybersecurity, privacy, professional accountability, and post-deployment monitoring. These risks are interdependent: a fluent answer may conceal uncertainty, a technically correct retrieval may be clinically outdated, and a high-performing model may fail when the patient population, workflow, device, prompt, or source data differ from those used during development.

The principal safety, validation, and governance requirements for LLM-enabled cardiovascular decision support are summarized in [Figure 3](#).

9.1. Data quality, bias, domain shift, and out-of-distribution performance

Clinical performance is bounded by the quality and representativeness of source data. Missing values, corrupted

physiological signals, inconsistent coding, device-dependent preprocessing, incomplete EHR histories, and temporal misalignment can generate downstream errors even when the model itself functions as designed. Dataset imbalance may further produce subgroup-specific performance gaps related to age, sex, ethnicity, language, socioeconomic context, comorbidity, geography, or healthcare setting. Therefore, aggregate accuracy should be complemented by prespecified subgroup analyses, calibration assessment, and testing across institutions, devices, and populations.

Domain shift and out-of-distribution inputs are particularly important in cardiovascular medicine because disease prevalence, acquisition protocols, imaging equipment, ECG devices, laboratory assays, documentation styles, and treatment pathways vary across institutions and over time. Models should therefore be evaluated for performance degradation when deployed outside their development environment. Multicenter external validation, prospective silent deployment, predefined failure thresholds, and continuous surveillance are preferable to assuming that retrospective internal performance will generalize to routine care.

9.2. Hallucinations, omission errors, prompt sensitivity, and automation bias

Hallucination remains a major safety concern, but clinically important failure is broader than fabrication alone. An LLM may provide a plausible but false statement, omit a critical contraindication or diagnostic alternative, overstate weak evidence, or produce different recommendations after clinically irrelevant changes in wording or prompt structure. In cardiovascular emergencies, omission errors may be as consequential as explicit factual errors because failure to surface time-critical information can delay appropriate treatment. Safety evaluation should therefore measure not only factual accuracy but also completeness, consistency across prompt variants, clinically significant error severity, and the model's ability to abstain when evidence is insufficient.

Human oversight does not automatically eliminate these risks. Automation bias may lead clinicians to accept fluent AI-generated recommendations despite conflicting clinical evidence, particularly under time pressure or high workload. Physician-in-the-loop workflows should therefore be designed around active verification rather than

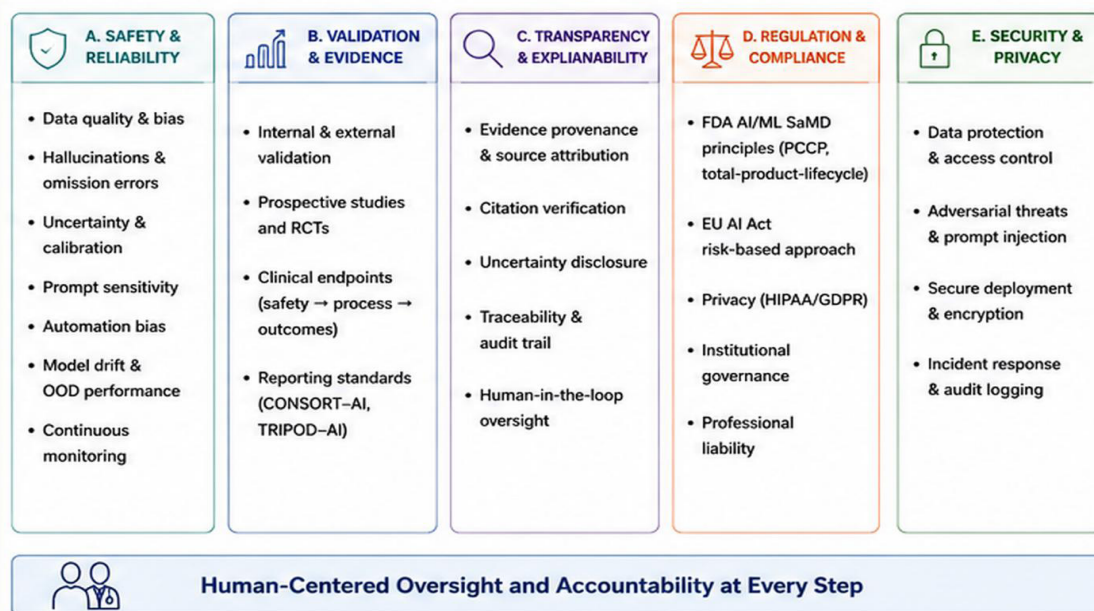


Figure 3. Safety, validation, and governance framework for LLM-enabled cardiovascular decision support. The schematic presents a framework summarizing major requirements for the responsible clinical implementation of LLM-enabled cardiovascular decision support, including data quality, uncertainty and calibration, hallucination and omission errors, evidence provenance, human oversight, prospective validation, privacy, cybersecurity, regulation, governance, and auditability. These safeguards apply across the lifecycle of clinical AI systems and are essential for accountable physician-supervised use. Schematic created by the author.

Abbreviations: AI: Artificial intelligence; FDA: US Food and Drug Administration; GDPR: General Data Protection Regulation; HIPAA: Health Insurance Portability and Accountability Act; ML: Machine learning; OOD: Out-of-distribution; PCCP: Predetermined change control plan; RCT: Randomized controlled trial.

passive approval. Interfaces should expose source evidence, unresolved conflicts, missing data, and uncertainty, and should avoid presenting probabilistic model outputs as authoritative conclusions.

9.3. Explainability, evidence provenance, and auditability

A fluent natural-language explanation should not be equated with a faithful representation of a model's internal reasoning. LLM-generated rationales may be persuasive while remaining post hoc, incomplete, or unsupported. For high-stakes cardiovascular decision support, operational explainability should therefore emphasize verifiable evidence provenance rather than rhetorical plausibility. Clinicians should be able to identify which patient data, specialist-model outputs, guideline passages, and external sources contributed to a recommendation and which elements remain uncertain or conflicting.

Trustworthy systems should provide source attribution, versioned guideline citations, timestamps, traceable retrieval records, uncertainty reporting, and auditable logs of model inputs, outputs, model version, prompts, tool calls, and clinician actions. Where calibrated probabilities are available from dedicated predictive models, these should be preserved rather than converted into unsupported verbal certainty by an orchestration LLM. Auditability is also essential for quality assurance, incident review, reproducibility, and medico-legal accountability.

9.4. Retrieval-augmented generation: Benefits and failure modes

Retrieval-augmented generation (RAG) can improve access to current guidelines and patient-specific evidence by supplying external documents at inference time, but it does not eliminate hallucination or guarantee adherence to clinical rules. Failure may occur because the retriever selects an irrelevant document, misses a key recommendation, retrieves an outdated version, combines conflicting guidelines, or provides context that the generator subsequently misinterprets. Citation generation itself must also be verified because a plausible citation does not establish that the cited source supports the generated claim.

Clinical RAG systems should therefore use curated and version-controlled knowledge bases, explicit source hierarchies, retrieval-quality evaluation, citation verification, conflict detection, and mechanisms for abstention when evidence is insufficient. Prompt injection and adversarial content within retrieved documents represent additional risks, particularly when external or untrusted sources are indexed. Retrieval logs and

document versions should be retained so that a clinical recommendation can be reconstructed after the event.

9.5. Prospective clinical validation and model monitoring

Randomized evidence remains limited and is not uniformly positive. In a multicenter randomized clinical vignette trial of 50 physicians, access to GPT-4 did not significantly improve diagnostic reasoning compared with conventional resources (median score 76% versus 74%; difference 2 percentage points, 95% CI -4 to 8; $p = 0.60$). By contrast, a subsequent prospective randomized trial of 92 physicians evaluating open-ended patient-care management tasks found that GPT-4 assistance improved rubric-based performance by 6.5 percentage points (95% CI 2.7 to 10.2; $p < 0.001$), although physicians using the LLM spent more time per case. These findings suggest that the benefit of LLM assistance is task-dependent and should not be inferred from benchmark performance alone.^{51,70}

Cardiology-specific randomized evidence is emerging but remains narrow. In 107 complex real-world cases involving suspected inherited cardiomyopathy, nine general cardiologists were randomized to assess cases with or without AMIE assistance. Blinded subspecialists preferred AMIE-assisted assessments in 46.7% of comparisons versus 32.7% for cardiologists alone ($p = 0.02$); clinically significant errors were lower with AMIE assistance (13.1% versus 24.3%; $p = 0.033$), and missing content was less frequent (17.8% versus 37.4%; $p = 0.0021$). However, the trial used retrospective patient data from a subspecialty setting and therefore does not establish improved cardiovascular outcomes, reduced hospitalization, or mortality.⁷

Prospective validation should test whether LLM assistance improves clinically meaningful processes or outcomes rather than merely reproducing examination-style answers. A hierarchical endpoint strategy may be useful: first, safety endpoints such as clinically significant error, omission, inappropriate treatment, and failure-to-escalate rates; second, process endpoints such as diagnostic accuracy, calibration, time to diagnosis, treatment appropriateness, guideline adherence, time to reperfusion, clinician workload, and decision time; and third, patient-centered outcomes such as readmission, major adverse cardiovascular events, mortality, quality of life, and patient comprehension. The appropriate endpoint set should be prespecified for the intended clinical use and compared with usual care or an appropriate specialist comparator.

Randomized or pragmatic trials are desirable for high-impact decision-support applications, supplemented by multicenter external validation and prospective silent-

mode evaluation before active deployment. Adaptive or frequently updated systems require additional safeguards because performance may change after model updates, retrieval-corpus changes, fine-tuning, or shifts in local practice. Version control, change documentation, revalidation after material modifications, drift detection, rollback procedures, and post-deployment surveillance should therefore form part of the clinical lifecycle.^{71,72,73}

9.6. Regulation, privacy, institutional governance, and professional liability

Regulatory concepts should be separated according to their function. In the United States, the Food and Drug Administration (FDA) regulates medical devices and device software when the intended use and product characteristics bring a system within the applicable device framework; AI-enabled device functions may also require lifecycle controls and, for planned modifications, a predetermined change control plan. In the European Union, the AI Act establishes horizontal, risk-based obligations for AI systems and does not replace medical-device requirements. Privacy and data-protection obligations, such as those arising under the General Data Protection Regulation (GDPR), constitute a separate legal layer. These frameworks should therefore not be presented as equivalent regulatory pathways.⁷⁴

Institutional governance is distinct again from formal regulation. Healthcare organizations require policies defining authorized use cases, data access, cybersecurity, procurement, local validation, model and prompt versioning, incident reporting, audit frequency, human verification, and procedures for suspension when performance deteriorates. Professional liability likewise remains a separate question: the availability of AI support does not remove the clinician's duty to exercise professional judgment, while responsibility may also involve institutions, developers, and manufacturers depending on system design, intended use, and jurisdiction.

Continuously re-evaluated or adaptive models create particular governance challenges because the system used tomorrow may not be functionally identical to the system originally validated. Change management should therefore be evidence-based, bounded, transparent, and linked to ongoing performance monitoring. Regulatory lifecycle approaches for AI-enabled device software increasingly emphasize predefined modification protocols, validation of changes, transparency to users, and mechanisms to detect and respond to performance deterioration.

9.7. Privacy, cybersecurity, and adversarial threats

Large language model (LLM)-enabled cardiovascular systems may access EHRs, imaging, laboratory results, genomic information, wearable data, and retrieved external knowledge. Security architecture should therefore address encryption, authentication, role-based access, data minimization, retention, audit trails, and protection against unauthorized disclosure. Where appropriate, privacy-preserving deployment strategies such as secure on-premises environments or federated approaches may reduce exposure, although they do not remove the need for governance and validation.

Adversarial threats extend beyond conventional data breaches. Prompt injection, malicious or contaminated retrieved documents, manipulated inputs, model extraction, and tool misuse may alter generated recommendations or expose sensitive information. Systems that can call external tools or retrieve documents should apply least-privilege access, source allow-lists where feasible, content validation, isolation of untrusted instructions, and continuous security testing.

9.8. Future multimodal cardiovascular intelligence

Future cardiovascular systems are likely to combine validated modality-specific algorithms, multimodal foundation models, longitudinal EHR data, biomedical sensors, and evidence retrieval within physician-supervised orchestration workflows. Their clinical value will depend less on whether a single model can process every modality than on whether the overall system can preserve data provenance, quantify uncertainty, detect failure, integrate conflicting evidence, and support timely human decisions. The integrative intelligence framework proposed in this review should therefore be tested as a clinical systems hypothesis rather than assumed to confer outcome benefit.

Future research should prioritize prospective multicenter studies, standardized reporting, subgroup and out-of-distribution evaluation, robust calibration, clinically meaningful endpoints, reproducible versioning, and transparent post-deployment surveillance. The objective is not autonomous cardiovascular decision-making but a traceable and accountable human-AI partnership in which computational integration strengthens, rather than obscures, clinical judgment.^{75,76,77}

The principal strengths, current limitations, and future directions of LLM-enabled integrative intelligence in cardiovascular medicine are summarized in [Table 4](#).

Table 4. Strengths, current limitations, and future directions of LLM-enabled integrative intelligence in cardiovascular medicine

Domain	Strengths	Current limitations	Future directions
Multimodal data integration	Simultaneous interpretation of EHRs, ECG, echocardiography, cardiac CT/MRI, laboratory biomarkers, and clinical documentation	Data heterogeneity, interoperability challenges, incomplete datasets	Standardized multimodal foundation models and seamless interoperability
Clinical reasoning	Context-aware synthesis of complex clinical information, differential diagnosis support, evidence integration	Hallucinations, limited explainability, dependence on input quality	Explainable AI, uncertainty estimation, physician-in-the-loop validation
Guideline interpretation	Rapid retrieval and summarization of contemporary ESC and ACC/AHA recommendations	Potential use of outdated information without retrieval augmentation	Real-time integration with continuously updated clinical guidelines
Clinical documentation	Automated generation of reports, discharge summaries, patient education materials, reduced administrative burden	Requires clinician review, possible factual inaccuracies	Fully integrated documentation workflows with human verification
Personalized decision support	Individualized risk stratification and treatment recommendations based on multimodal data	Limited prospective validation, uncertain generalizability across populations	Precision cardiovascular medicine supported by longitudinal learning systems
Clinical workflow efficiency	Improved information accessibility, multidisciplinary communication, reduced cognitive workload	Integration into hospital information systems remains incomplete	Intelligent workflow orchestration across the entire care pathway
Ethical and legal aspects	Improved transparency through explainable outputs, support for shared decision-making	Accountability, bias, privacy protection, regulatory uncertainty	Robust governance frameworks, international standards, AI regulation
Future translational potential	Human-AI collaborative intelligence, continuous learning, research acceleration	Need for large prospective clinical trials and regulatory approval	Learning healthcare systems integrating multimodal foundation models and real-world evidence

Abbreviations: ACC: American College of Cardiology; AHA: American Heart Association; AI: Artificial intelligence; CT: Computed tomography; ECG: Electrocardiography; EHR: Electronic health record; ESC: European Society of Cardiology; LLM: large language model; MRI: Magnetic resonance imaging.

10. Conclusions

Large language models (LLMs) are increasingly relevant to cardiovascular informatics, but the current evidence does not support treating a conventional text-based LLM as a universal multimodal cardiovascular model. Dedicated ECG and imaging algorithms, multimodal foundation models, and text-based LLMs have distinct technical roles and should be evaluated according to their intended inputs and clinical tasks.

The integrative intelligence concept proposed in this review therefore refers to a physician-supervised orchestration framework in which validated modality-specific outputs, clinical text, EHR data, biomarkers, and retrieved evidence can be synthesized into a traceable decision-support representation. Its proposed contribution is organizational and translational rather than architectural, and the complete framework has not yet been prospectively validated as an end-to-end intervention.

Early randomized and comparative studies suggest that LLM assistance may improve selected reasoning,

documentation, or decision-support tasks, but benefits are task-dependent and evidence for improved cardiovascular outcomes remains insufficient. Future evaluation should prioritize multicenter external validation, prospective and pragmatic trials, clinically meaningful safety and outcome endpoints, calibration and subgroup analyses, model-version control, evidence provenance, auditability, cybersecurity, and post-deployment monitoring.

Accordingly, the near-term objective should not be autonomous cardiovascular decision-making. A more defensible goal is a transparent human-AI partnership in which specialized analytical systems and LLM-enabled orchestration support, but do not replace, accountable clinical judgment.

Acknowledgments

None.

Funding

None.

Conflict of interest

The author declares no conflicts of interest related to this work.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable. This article is a narrative review based exclusively on previously published literature. No human participants, animals, or identifiable personal data were involved. Therefore, ethics committee approval and informed consent were not required.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

- Boonstra MJ, Weissenbacher D, Moore JH, Gonzalez-Hernandez G, Asselbergs FW. Artificial intelligence: revolutionizing cardiology with large language models. *Eur Heart J*. 2024;45(5):332-345.
doi: 10.1093/eurheartj/ehad838
- Wehbe RM. Charting the future of cardiology with large language model artificial intelligence. *Nat Rev Cardiol*. 2025;22:143-144.
doi: 10.1038/s41569-024-01105-y
- Meng X, Yan X, Zhang K, *et al*. The application of large language models in medicine: a scoping review. *iScience*. 2024;27(5):109713.
doi: 10.1016/j.isci.2024.109713
- Gendler M, Nadkarni GN, Sudri K, *et al*. Large language models in cardiology: systematic review. *JMIR Cardio*. 2026;10:e76734.
doi: 10.2196/76734
- Agrawal M, Hegselmann S, Lang H, Kim Y, Sontag D. Large language models are few-shot clinical information extractors. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics; 2022:1998-2022.
doi: 10.18653/v1/2022.emnlp-main.130
- Zakka C, Shad R, Chaurasia A, *et al*. Almanac-retrieval-augmented language models for clinical medicine. *NEJM AI*. 2024;1(2):AIoa2300068.
doi: 10.1056/AIoa2300068
- O'Sullivan JW, Palepu A, Saab K, *et al*. A large language model for complex cardiology care. *Nat Med*. 2026;32(2):616-623.
doi: 10.1038/s41591-025-04190-9
- Ferreira Santos J, Ladeiras-Lopes R, Leite F, Dorés H. Applications of large language models in cardiovascular disease: a systematic review. *Eur Heart J Digit Health*. 2025;6(4):540-553.
doi: 10.1093/ehjdh/ztaf028
- Khera R, Oikonomou EK, Nadkarni GN, *et al*. Transforming cardiovascular care with artificial intelligence: from discovery to practice: JACC state-of-the-art review. *J Am Coll Cardiol*. 2024;84(1):97-114.
doi: 10.1016/j.jacc.2024.05.003
- Kodera S, Takeda N. From DNA to drug discovery: AI models for cardiovascular precision medicine. *J Cardiol*. 2026.
doi: 10.1016/j.jjcc.2026.06.007
- Johnson KW, Torres Soto J, Glicksberg BS, *et al*. Artificial intelligence in cardiology. *J Am Coll Cardiol*. 2018;71(23):2668-2679.
doi: 10.1016/j.jacc.2018.03.521
- Makimoto H, Kohro T. Adopting artificial intelligence in cardiovascular medicine: a scoping review. *Hypertens Res*. 2024;47(3):685-699.
doi: 10.1038/s41440-023-01469-7
- Moor M, Banerjee O, Abad ZSH, *et al*. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259-265.
doi: 10.1038/s41586-023-05881-4
- Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-1940.
doi: 10.1038/s41591-023-02448-8
- Omar M, Nadkarni GN, Klang E, Glicksberg BS. Large language models in medicine: a review of current clinical trials across healthcare applications. *PLoS Digit Health*. 2024;3(11):e0000662.
doi: 10.1371/journal.pdig.0000662
- Hager P, Jungmann F, Holland R, *et al*. Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nat Med*. 2024;30(9):2613-2622.
doi: 10.1038/s41591-024-03097-1
- McCoy LG, Manrai AK, Rodman A. Large language models and the degradation of the medical record. *N Engl J Med*. 2024;391(17):1561-1564.
doi: 10.1056/NEJMp2405999
- Kung TH, Cheatham M, Medenilla A, *et al*. Performance

- of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digit Health*. 2023;2(2):e0000198.
doi: 10.1371/journal.pdig.0000198
19. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023;388(13):1233-1239.
doi: 10.1056/NEJMSr2214184
20. Yang XY, Li YM, Wang JY, Jia YH, Yi Z, Chen M. Utilizing multimodal artificial intelligence to advance cardiovascular diseases. *Precis Clin Med*. 2025;8(3):pbaf016.
doi: 10.1093/pcmedi/pbaf016
21. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2015;16(3):233-271.
doi: 10.1093/ehjci/jev014
22. Voigt JU, Cvijic M. 2- and 3-dimensional myocardial strain in cardiac health and disease. *JACC Cardiovasc Imaging*. 2019;12(9):1849-1863.
doi: 10.1016/j.jcmg.2019.01.044
23. Li J, Li Y, Sun Z, et al. Exploring multimodal large language models on transthoracic echocardiogram tasks for cardiovascular decision support. *J Biomed Inform*. 2025;171:104930.
doi: 10.1016/j.jbi.2025.104930
24. Chao CJ, Banerjee I, Arsanjani R, et al. Evaluating large language models in echocardiography reporting: opportunities and challenges. *Eur Heart J Digit Health*. 2025;6(3):326-339.
doi: 10.1093/ehjdh/ztae086
25. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25:70-74.
doi: 10.1038/s41591-018-0240-2
26. Ribeiro AH, Ribeiro MH, Paixao GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11(1):1760.
doi: 10.1038/s41467-020-15432-4
27. Lee H, Yoo S, Kim J, Cho Y, Suh D, Lee K. Comparative diagnostic performance of a multimodal large language model versus a dedicated electrocardiogram AI in detecting myocardial infarction from electrocardiogram images: comparative study. *JMIR AI*. 2025;4:e75910.
doi: 10.2196/75910
28. Rouhi AD, Menon SV, Ghanem YK, Han JJ. Concordance of large language model recommendations with multidisciplinary heart team decisions in coronary revascularization and aortic valve intervention: A systematic review and pooled analysis. *Cardiol Ther*. 2026.
doi: 10.1007/s40119-026-00453-9
29. Gulati M, Levy PD, Mukherjee D, et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain. *Circulation*. 2021;144(22):e368-e454.
doi: 10.1161/CIR.0000000000001029
30. Hundley WG, Bluemke DA, Finn JP, et al. ACCF/ACR/AHA/NASCI/SCMR expert consensus document on cardiovascular magnetic resonance. *J Am Coll Cardiol*. 2010;55:2614-2662.
doi: 10.1161/CIR.0b013e3181d44a8f
31. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction. *Circulation*. 2018;138(20):e618-e651.
doi: 10.1161/CIR.0000000000000617
32. Tsutsui H, Albert NM, Coats AJS, et al. Natriuretic Peptides: Role in the diagnosis and management of heart failure: A Scientific statement from the Heart Failure Association of the European Society of Cardiology, Heart Failure Society of America and Japanese Heart Failure Society. *Eur J Heart Fail*. 2023;25(5):616-631.
doi: 10.1002/ehf.2848
33. Xu D, Cunningham JW. Harnessing large language models for chart review. *J Am Heart Assoc*. 2025;14(7):e041581.
doi: 10.1161/JAHA.125.041581
34. Jain SS, Elias P, Poterucha TJ, et al. Artificial intelligence in cardiovascular care-part 2: applications: JACC review topic of the week. *J Am Coll Cardiol*. 2024;83(24):2487-2496.
doi: 10.1016/j.jacc.2024.03.401
35. Ke YH, Jin L, Elangovan K, et al. Retrieval augmented generation for 10 large language models and its generalizability in assessing medical fitness. *npj Digit Med*. 2025;8(1):187.
doi: 10.1038/s41746-025-01519-z
36. Templin T, Fort S, Padmanabham P, et al. Framework for bias evaluation in large language models in healthcare settings. *npj Digit Med*. 2025;8(1):414.
doi: 10.1038/s41746-025-01786-w
37. Radke RM, Diller GP, Reddy RG, Shivaram P, Danford DA, Kutty S. A multi-query, multimodal, receiver-augmented solution to extract contemporary cardiology guideline information using large language models. *Eur Heart J Digit Health*. 2025.
doi: 10.1093/ehjdh/ztaf111

38. Cheema B, Pandit J. AI and heart failure: present state and future with multimodal large language models. *JACC Adv.* 2024;3(9):101029.
doi: 10.1016/j.jacadv.2024.101029
39. Nolin-Lapalme A, *et al.* Maximising large language model utility in cardiovascular care: a practical guide. *Can J Cardiol.* 2024;40(10):1774-1787.
doi: 10.1016/j.cjca.2024.05.024
40. Perpetua EM, Palmer R, Le VT, *et al.* JACC: Advances expert panel perspective: shared decision-making in multidisciplinary team-based cardiovascular care. *JACC Adv.* 2024;3(7):100981.
doi: 10.1016/j.jacadv.2024.100981
41. Antiperovitch P, Liu I, Mokhtar AT, Tang A. Evaluating large language models in cardiovascular antithrombotic care: performance, accuracy, and implications for clinical practice. *Can J Cardiol.* 2025;41(8):1584-1591.
doi: 10.1016/j.cjca.2025.04.008
42. Byrne RA, Rossello X, Coughlan JJ, *et al.* 2023 ESC guidelines for the management of acute coronary syndromes. *Eur Heart J.* 2023;44(38):3720-3826.
doi: 10.1093/eurheartj/ehad191
43. McDonagh TA, Metra M, Adamo M, *et al.* 2023 focused update of the 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2023;44(37):3627-3639.
doi: 10.1093/eurheartj/ehad195
44. Heidenreich PA, Bozkurt B, Aguilar D, *et al.* 2022 AHA/ACC/HFSA guideline for the management of heart failure. *Circulation.* 2022;145(18):e895-e1032.
doi: 10.1161/CIR.0000000000001063
45. Joglar JA, Chung MK, Armbruster AL, *et al.* 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation. *Circulation.* 2024;149(1):e1-e156.
doi: 10.1161/CIR.0000000000001193
46. Lin C, Lan Y, Zeng Z, *et al.* Evaluation of large language models in percutaneous coronary intervention decision-making. *Front Cardiovasc Med.* 2026;13:1690716.
doi: 10.3389/fcvm.2026.1690716
47. Vahanian A, Beyersdorf F, Praz F, *et al.* 2021 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J.* 2022;43(7):561-632.
doi: 10.1093/eurheartj/ehab395
48. Zhu W, Peng J, Yan Z, Chen Y, Xu J, Zhang L. Designing and evaluating large language model-enabled clinical decision support for heart failure: a modular and risk-tiered framework. *Front Digit Health.* 2026;8:1730457.
doi: 10.3389/fdgh.2026.1730457
49. Lyon AR, Lopez-Fernandez T, Couch LS, *et al.* 2022 ESC guidelines on cardio-oncology. *Eur Heart J.* 2022;43(41):4229-4361.
doi: 10.1093/eurheartj/ehac244
50. Herrmann J, Lenihan D, Armenian S, *et al.* Defining cardiovascular toxicities of cancer therapies: an International Cardio-Oncology Society consensus statement. *Eur Heart J.* 2022;43(4):280-299.
doi: 10.1093/eurheartj/ehab674
51. Goh E, Gallo RJ, Strong E, *et al.* GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. *Nat Med.* 2025;31(4):1233-1238.
doi: 10.1038/s41591-024-03456-y
52. Wang G, Zhang K, Jiang J, *et al.* Human-large language model collaboration in clinical medicine: a systematic review and meta-analysis. *npj Digit Med.* 2026;9(1):195.
doi: 10.1038/s41746-026-02382-2
53. Ong KT, Seo J, Kim H, *et al.* Success and failure of human-AI collaboration in clinical reasoning: an experimental study on challenging real-world cases. *Int J Med Inform.* 2026;211:106342.
doi: 10.1016/j.ijmedinf.2026.106342
54. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44-56.
doi: 10.1038/s41591-018-0300-7
55. World Health Organization. Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models. World Health Organization; 2024. <https://www.who.int/publications/i/item/9789240084759>
56. Wu J, Wu X, Zheng Y, Yang J. Clinical pathway-aware large language models for reliable and transparent medical dialogue. *J Biomed Inform.* 2025;172:104942.
doi: 10.1016/j.jbi.2025.104942
57. Elwyn G, Frosch DL, Kobrin S. Implementing shared decision-making: consider all the consequences. *Implement Sci.* 2016;11:114.
doi: 10.1186/s13012-016-0480-9
58. Jiang K, Adhikari J, Bernard GR. Interpreting ECG Images with Multimodal Large Language Models. In: *Studies in Health Technology and Informatics*. Amsterdam: IOS Press; 2026.
doi: 10.3233/shti260321
59. Jia YY, Pang LY, Bi MM, Yang XL, Song JP. Dependability of large language models in cardiovascular medicine: a scoping review. *J Cardiothorac Vasc Anesth.* 2025;39(12):3534-3540.
doi: 10.1053/j.jvca.2025.07.026

60. Mahajan A, Obermeyer Z, Daneshjou R, Lester J, Powell D. Cognitive bias in clinical large language models. *npj Digit Med*. 2025;8:428.
doi: 10.1038/s41746-025-01790-0
61. Yang Y, Liu X, Jin Q, Huang F, Lu Z, *et al*. Unmasking and quantifying racial bias of large language models in medical report generation. *Commun Med*. 2024;4:176.
doi: 10.1038/s43856-024-00601-z
62. Zhong X, Li S, Chen Z, Ge L, Yu D, Wang S, *et al*. Considerations for patient privacy of large language models in health care: scoping review. *J Med Internet Res*. 2025;27:e76571.
doi: 10.2196/76571
63. Omar M, Sorin V, Collins JD, *et al*. Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Commun Med*. 2025;5(1):330.
doi: 10.1038/s43856-025-01021-3
64. Asgari E, Montana-Brown N, Dubois M, *et al*. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *npj Digit Med*. 2025;8(1):274.
doi: 10.1038/s41746-025-01670-7
65. Amugongo LM, Mascheroni P, Brooks S, Doering S, Seidel J. Retrieval augmented generation for large language models in healthcare: a systematic review. *PLoS Digit Health*. 2025;4(6):e0000877.
doi: 10.1371/journal.pdig.0000877
66. Shool S, Adimi S, Saboori Amleshi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model evaluations in clinical medicine. *BMC Med Inform Decis Mak*. 2025;25(1):117.
doi: 10.1186/s12911-025-02954-4
67. Johri S, Jeong J, Tran BA, *et al*. An evaluation framework for clinical use of large language models in patient interaction tasks. *Nat Med*. 2025;31(1):77-86.
doi: 10.1038/s41591-024-03328-5
68. Bibbò L, Laganà F, Pullano SA, Angiulli G. Multimodal EEG-EMG and FEM-based adaptive control of passive upper-limb exoskeletons. *Sensors*. 2026;26(12):3924.
doi: 10.3390/s26123924
69. Qu C, Zhang X, Lu Y, Wang Y, Su C. MAF-Net: multimodal cross-attention-based fusion network for cardiovascular disease classification. *PLoS ONE*. 2026;21(4):e0345238.
doi: 10.1371/journal.pone.0345238
70. Goh E, Gallo R, Hom J, *et al*. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw Open*. 2024;7(10):e2440969.
doi: 10.1001/jamanetworkopen.2024.40969
71. Wiens J, Saria S, Sendak M, *et al*. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;25(9):1337-1340.
doi: 10.1038/s41591-019-0548-6
72. Gallifant J, Afshar M, Ameen S, *et al*. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat Med*. 2025;31(1):60-69.
doi: 10.1038/s41591-024-03425-5
73. Kagiya N, Tokodi M, Hathaway QA, *et al*. PRIME 2.0: proposed requirements for cardiovascular imaging-related multimodal-AI evaluation: an updated checklist. *JACC Cardiovasc Imaging*. 2026;19(2):225-251.
doi: 10.1016/j.jcmg.2025.08.004
74. European Parliament, Council of the European Union. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Off J Eur Union*. 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj?locale=en>
75. 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
76. Ferreira Santos J, Doros H. Large language models in cardiovascular prevention: a narrative review and governance framework. *Diagnostics*. 2026;16(3):390.
doi: 10.3390/diagnostics16030390
77. Tran M, Balasooriya C, Jonnagaddala J, *et al*. Situating governance and regulatory concerns for generative artificial intelligence and large language models in medical education. *npj Digit Med*. 2025;8(1):315.
doi: 10.1038/s41746-025-01721-z

