Echo-CoPilot: A Multiple-Perspective Agentic Framework for Reliable Echocardiography Interpretation

Moein Heidari[1†], Ali Mehrabian[1], Mohammad Amin Roohi[1], River Jiang[1], Wenjin Chen[2], David J. Foran[2], Jasmine Grewal[1], and Ilker Hacihaliloglu[1]

[1] The University of British Columbia, Vancouver, BC, Canada
[2] Rutgers Cancer Institute, New Brunswick, NJ, USA
moein.heidari@ubc.ca

Abstract. Echocardiography interpretation requires integrating multi-view temporal evidence with quantitative measurements and guideline-grounded reasoning, yet existing foundation-model pipelines largely solve isolated subtasks and fail when tool outputs are noisy, or values fall near clinical cutoffs. We propose Echo-CoPilot, an end-to-end agentic framework that combines a multi-perspective workflow with knowledge-graph–guided measurement selection. Echo-CoPilot runs three independent ReAct-style agents, structural, pathological, and quantitative, that invoke specialized echocardiography tools to extract parameters while querying EchoKG to determine which measurements are required for the clinical question and which should be avoided. A self-contrast language model then compares the evidence-grounded perspectives, generates a discrepancy checklist, and re-queries EchoKG to apply the appropriate guideline thresholds and resolve conflicts, reducing hallucinated measurement selection and borderline flip-flops. On MIMICEchoQA, Echo-CoPilot provides higher accuracy compared to SOTA baselines and, under a stochasticity stress test, achieves higher reliability through more consistent conclusions and fewer answer changes across repeated runs. Our code is publicly available at GitHub.

Keywords: Echocardiography · knowledge-graph · Language models.

1 Introduction

Echocardiography interpretation demands clinically reliable decision-making by integrating multi-view temporal evidence, quantitative measurements, and strict alignment with clinical guidelines [10]. While deep learning has achieved strong performance on isolated perceptual tasks such as view classification, segmentation, and functional quantification, full-study interpretation remains a bottleneck. Large language models (LLMs) and agentic systems offer a promising paradigm by treating interpretation as a procedure, dynamically invoking specialized perceptual tools to synthesize a final answer [21]. However, reliance on

† Corresponding author

a single reasoning trajectory may reduce robustness in clinical settings, particularly near severity thresholds, where small variations in measurements can alter the diagnostic conclusion [6]. This sensitivity is further amplified in echocardiography, where conclusions depend on specific views and guideline criteria, and where tool outputs may be noisy or occasionally inconsistent across views [18]. To address this, we propose Echo-CoPilot, a multi-perspective agentic framework for echocardiography interpretation that improves both performance and reliability by explicitly contrasting complementary clinical reasoning pathways. Our main contributions are: ❶ ***EchoKG***: We introduce the Echocardiography knowledge-graph (EchoKG), a clinical knowledge-graph derived from consensus echocardiography guidelines, capturing diagnostic dependencies and formalizing measurement selection and avoidance rules as structured inference-time constraints. ❷ ***Multiple-Perspective Framework***: We propose a multi-perspective procedure. Echo-CoPilot instantiates three parallel ReAct agents, Structural, Pathological, and Quantitative, that independently reason over the exam using a shared tool cache that reuses tool output to improve efficiency. A self-contrast LLM then aggregates these diverse trajectories, generating a discrepancy checklist to resolve conflicts while ensuring the final decision remains constrained by EchoKG. ❸ ***Performance Evaluation***: We achieve state-of-the-art (SOTA) accuracy on the MIMICEchoQA benchmark and empirically demonstrate that our agentic framework improves diagnostic stability and reduces hallucination and reasoning variance compared to single-trajectory baselines.

2 Related Work

Vision Foundation Models in Echocardiography. While early methods focused on narrow, task-specific supervision [11], recent efforts leverage scalable foundation models. Vision models like EchoApex [2] and MedSAM2 [9] provide robust representation and promptable segmentation, while vision-language models (VLMs) like PanEcho [7] and EchoPrime [20] map entire video studies to clinical descriptors or reports. However, these end-to-end models operate as "black boxes" [13], lacking the multi-step reasoning needed to decompose complex clinical queries or audit their logic against medical guidelines.

Agentic Frameworks in Medical Imaging. Tool augmented LLM agents aim to increase transparency by decomposing a task and invoking specialist tools [21]. Radiology agents like MedRAX [5] decompose X-ray interpretation into verifiable steps, but transferring this paradigm to echocardiography requires handling complex temporal dynamics and view-dependent feasibility. While concurrent works like EchoAgent [4] explore tool-assisted measurements, their single-trajectory reasoning cannot resolve the cross-tool and cross-view disagreements that are intrinsic to temporal echocardiographic evidence.

Constrained and verifiable reasoning. Recent work improves reliability through self-critique and external grounding, including reflective loops and structured knowledge integration [8,12]. Echo-CoPilot builds on these directions but targets echocardiography-specific failure modes. We use our proposed EchoKG

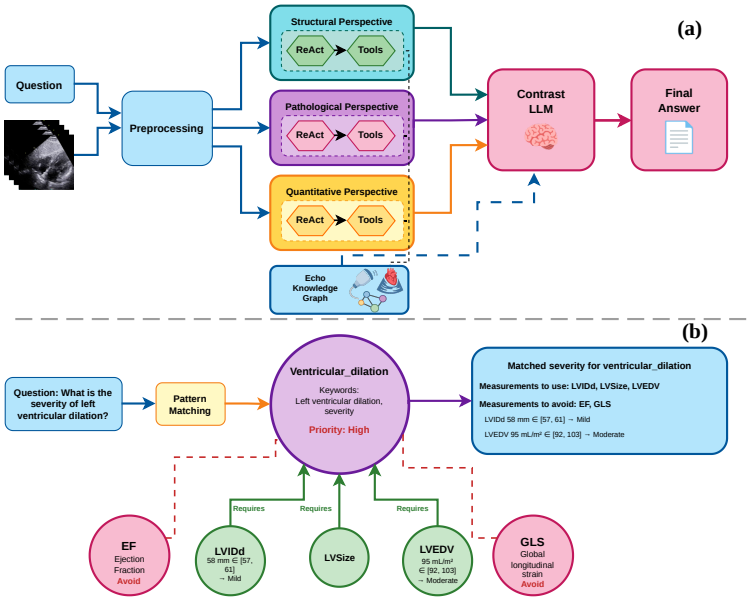

Fig. 1: Overview of Echo-CoPilot. Panel (a) shows the three perspective ReAct-style agents and the contrast module. Panel (b) illustrates EchoKG measurement selection/avoidance and threshold-based grading of ventricular dilation severity.

to enforce measurement selection guided by the question intent and avoidance at inference time, and propose a multiple-perspective procedure [22] to surface inconsistencies and synthesize a checklist-guided final answer grounded in tool evidence.

3 Method

Echo-CoPilot, an LLM-driven agentic system, decomposes a query into actionable steps, selectively invokes specialized tools, integrates intermediate findings, and synthesizes a transparent, clinically aligned final assessment. As illustrated in Figure 1, our proposed framework employs a two-stage architecture for robust echo interpretation: (1) multiple-perspective generation and (2) self-contrast analysis. Our proposed EchoKG is shared between these two stages to leverage an additional source of knowledge and reduce hallucinations.

Multi-Perspective Generation. Inspired by [22], given a question q and video v, we generate three independent perspectives using specialized system prompts:

- **Structural Perspective** (P_1): Focuses on anatomical structures, morphology, and chamber dimensions.
- **Pathological Perspective** (P_2): Emphasizes disease patterns and clinical indicators.

– **Quantitative Perspective** (P_3): Prioritizes numerical measurements and quantitative thresholds.

Each perspective executes an independent ReAct loop [21], autonomously selecting its specific tools (e.g., echo measurement prediction, echo disease prediction) and producing a final answer. The EchoKG is also shared between all agents to use the clinical knowledge and thresholds. We design a shared deterministic tool cache in which identical tool calls return the same cached output, eliminating redundancy while preserving consistency. The multi-perspective approach reduces hallucination by generating independent analyses that can cross-validate each other, catching errors that a single perspective might miss.
Echo-CoPilot is equipped with a set of domain-specific echocardiography tools, each implemented as a callable module with a structured input schema and standardized output format. In our implementation, the core tool set includes the EchoPrime model for view classification and measurement prediction [20], PanEcho for disease and finding prediction [7], report style synthesis [20], and our proposed EchoKG. Crucially, the reasoning agents are LLM-driven, while perception is delegated to dedicated vision–language tools.

Self-Contrast Mechanism. The contrast phase employs a dedicated contrast agent that analyzes the three perspectives $P = \{P_1, P_2, P_3\}$ and generates a structured discrepancy checklist. The contrast prompt includes: (1) all perspective responses, (2) aggregated measurements from tool outputs, (3) EchoKG guidance and clinical thresholds retrieved via EchoKG(q), and (4) validation instructions requiring the model to produce an explicit discrepancy checklist before its final answer. The contrast LLM first generates a structured discrepancy checklist by validating whether perspectives followed EchoKG guidance, identifying severity classification discrepancies, comparing quantitative measurement consistency, and assessing tool reliability. The contrast LLM then synthesizes a final answer that addresses all checklist items, resolves discrepancies, and uses EchoKG thresholds to determine severity classifications. The final answer integrates evidence from all perspectives while giving higher weight to perspectives that followed EchoKG guidance. The contrast phase explicitly validates measurement selection against EchoKG guidance, giving higher weight to perspectives that followed medical principles rather than simple majority voting.

Echocardiography knowledge-graph. To enforce clinically sound reasoning and prevent metric hallucination, we propose the EchoKG. As illustrated in Figure 1 (Panel (b)), EchoKG is a directed graph $G = (V, E)$ that uses guideline logic as queryable constraints via three node types: (1) **Structure nodes** (V_s) representing anatomical components, (2) **Measurement nodes** (V_m) defining standard metrics (e.g., LVEDV, EF), and (3) **Pattern nodes** (V_p) comprising 16 distinct clinical query templates (e.g., cavity dilation, valvular regurgitation). EchoKG governs deterministic measurement selection by mapping Pattern nodes to Measurement nodes via strict *requires* and *avoid* directed edges. The graph topology and mapping rules are meticulously distilled from major clinical guide-

Algorithm 1 Proposed Echo-CoPilot Agentic Framework

Input: Question q, video v, tools T, perspective prompts $\{\text{prompt}_i\}_{i=1}^{3}$
Output: Final answer A_{final}

$C \leftarrow \emptyset$ ▷ Shared cache for deterministic tool calls
$g \leftarrow \textsc{EchoKG}(q)$ ▷ Measurement selection and avoidance rules available to all stages
for $i \in \{1, 2, 3\}$ **do**
    $P_i \leftarrow \textsc{ReAct}(q, v, \text{prompt}_i, T; C, g)$ ▷ P_i includes answer and tool evidence, guided by g
end for
$m \leftarrow \textsc{Aggregate}(\{P_i\}_{i=1}^{3})$ ▷ Measurements from tool outputs
$\mathcal{L} \leftarrow \textsc{ContrastLLM}(\{P_i\}, m, g)$ ▷ Discrepancy checklist with EchoKG constraints
$A_{\text{final}} \leftarrow \textsc{Refine}(\{P_i\}, \mathcal{L}, m, g)$ ▷ Final constrained resolution
return A_{final}

lines (e.g., ASE, EACVI [10,19,18][3]), serving as a guardrail for the agent. Given question q, EchoKG(q) executes: (1) **Pattern matching**: Patterns are evaluated by priority (1=most specific, 3=most general), ensuring specific patterns (e.g., atrial enlargement) match before general ones (e.g., cavity dilation), reducing false positives. For pattern p with keywords K, context K_c, and exclusions K_e, p matches if $K \cap q \neq \emptyset$, $K_c \cap q \neq \emptyset$ (if K_c exists), and $K_e \cap q = \emptyset$. (2) **Structure disambiguation**: If p supports multiple structures (stored as metadata), detect the specific structure from q via keyword matching. (3) **Graph traversal**: Follow "requires" edges from p to get M_{req} and "avoid" edges to get M_{avoid}. Unlike traditional knowledge graphs with only positive edges, EchoKG employs dual-edge relationships with explicit "avoid" edges providing negative guidance (e.g., avoid EF for cavity size questions), preventing common measurement selection errors. (4) **Threshold retrieval**: Extract clinical thresholds (e.g., normal ranges, severity classifications) from measurement node attributes. EchoKG stores clinical thresholds in measurement nodes to provide guideline-based reference values and reduce hallucinated ranges. It is queried during perspective generation and contrast analysis with a shared cache for consistent guidance. The proposed framework is summarized in Algorithm 1.

4 Experiments

Dataset, Implementation Details, and Evaluation Metrics. We evaluate Echo-CoPilot on MIMICEchoQA [17], which contains 622 transthoracic echocardiogram videos paired with multiple-choice questions (four options: A–D) spanning 38 standard views. The questions cover core interpretation tasks, including ventricular systolic function, chamber size, valvular stenosis, and pericardial effusion. As this benchmark is strictly reserved for zero-shot evaluation, our agentic

[3] Incorporated materials are utilized strictly for knowledge construction and inference-time retrieval rather than model training, in full compliance with their terms of use.

Table 1: Accuracy (Acc.) of general-purpose and biomed-specialized VLMs on the MIMICEchoQA benchmark. Models marked with * are taken directly from OpenBiomedVid [17,16].

Model	Acc.	Model	Acc.
Video-ChatGPT*	31.7	Video-LLaVA*	32.0
Phi-3.5-Vision-Instruct*	41.1	Phi-4-Multimodal-Instruct*	37.8
InternVideo2.5-Chat-8B*	40.3	Qwen2-VL-7B-Instruct*	37.9
Qwen2.5-VL-7B-Instruct*	34.0	Qwen2 VL 72B Instruct*	37.5
Qwen2.5-VL-72B-Instruct*	34.2	Gemini-2.0-Flash*	38.4
GPT-4o*	41.6	GPT-o4-mini*	43.9
Qwen2-VL-2B-Biomed*	42.0	Qwen2-VL-7B-Biomed*	49.0
MedGemma-4B	33.4	LLaVA-Med	32.5
Echo-CoPilot	**54.0 ± 0.7**		

framework had no exposure to it during training; we report accuracy on the entire held-out dataset [17]. Echo-CoPilot uses open-source gpt-oss-120b [1] as the LLM and invokes tools through structured function calls implemented with the LangChain/LangGraph framework. Videos are preprocessed by frame sampling and resolution normalization to ensure stable cross-tool compatibility. All experiments are run on a single NVIDIA A100 (80GB) GPU. Other implementation choices include the 3-perspective prompts and LLM temperatures (contrast=0.3, perspective=0.7). For image-only baselines such as MedGemma 4B [14], we follow prior practice and evaluate on key frames extracted from each clip, since these models are not designed to ingest full video. For the specific analysis of system stability and agentic reasoning consistency, we created a stratified subset of 50 challenging question-answer pairs covering diverse categories (structural, functional, and disease-related queries). The maximum reasoning steps for the ReAct loop were set to 10. To validate the efficacy and robustness of Echo-CoPilot, our experiments focus on two key dimensions: (1) diagnostic accuracy compared to SOTA VLMs, and (2) response stability and consistency, demonstrating the specific value of the proposed multiple-perspective mechanism.

4.1 Diagnostic Accuracy

Table 1 reports accuracy on MIMICEchoQA. Echo-CoPilot achieves the best overall performance, outperforming both proprietary and open-source multimodal baselines. The gains are largest on questions that require quantitative reasoning or integrating multiple cues, such as grading systolic dysfunction from ejection fraction, assessing hypertrophy severity, or stratifying pericardial effusion. In these cases, vision-only models often rely on appearance-driven heuristics and are more likely to flip labels near guideline cutoffs, whereas Echo-CoPilot defers to exam-specific measurements and disease cues produced by its tools and

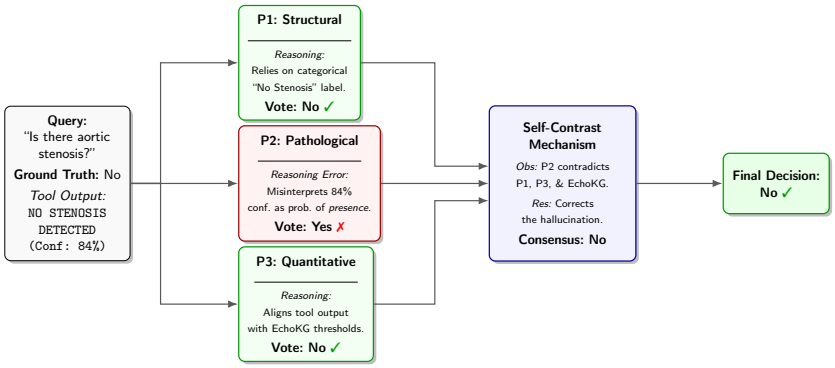

Fig. 2: Multiple-Perspective Error Correction. When the deterministic tool output is misinterpreted by a single reasoning perspective (e.g., P2), the self-contrast mechanism successfully isolates and corrects the logic error.

then applies guideline-grounded interpretation through EchoKG. Figure 2 shows a representative failure case where the value of the self-contrast (SC) mechanism is clear. For the query "Is there aortic stenosis?" (ground truth: No), one perspective incorrectly treated the tool confidence (84%) as the disease probability and voted Yes, while the other two perspectives correctly voted No based on the categorical tool output ("No stenosis detected"). The SC layer detected this conflict, used EchoKG to prioritize the categorical label, and recovered the correct diagnosis.

Table 2: Stability Comparison across Testing Modes: consistency over 10 runs for N=50 questions. Best results are in **bold.**

Metric	LLM	LLM+Tools	ReAct Agent	Echo-CoPilot
Stability Rate ↑	40.0%	26.0%	70.0%	**72.0%**
Avg Unique Answers ↓	1.86	2.00	1.30	**1.22**
Avg Changes ↓	2.10	2.58	1.14	**0.96**

4.2 Stability and Consistency Analysis, Ablation Study.

Clinical deployment demands reproducible decisions under stochastic decoding and noisy tool outputs [15]. We evaluate stability across 10 runs on 50 questions (500 total inferences) under four configurations: LLM, LLM+Tools (naive function calling), ReAct Agent, and Echo-CoPilot. Following [3], we measure *Stability Rate* (fraction of identical predictions across 10 runs), *Avg Unique Answers*, and *Avg Changes* (mean number of answer flips). Table 2 shows that naive tool access severely degrades stability (26%) by amplifying uncontextualized tool noise.

Table 3: Left: Error concentration on three frequent failure groups, comparing Qwen2-VL-7B-Biomed [16] with Echo-CoPilot. Right: Ablation study on each component of Echo-CoPilot: EchoKG, ReAct loop, and self-contrast (SC) mechanism. Best results are in **bold**.

Error Analysis (err.%)			**Ablation Study**					
Group	**Qwen**	**Ours**	**Variant**	**Tools**	**EchoKG**	**ReAct**	**SC**	**Acc.**
Doppler views	83.0	**59.3**	LLM	○	○	○	○	46.9
PSAX vessels	71.0	**59.6**	LLM+Tools	●	●	○	○	49.7
Left atrium	68.4	**60.5**	ReAct Agent	●	●	●	○	52.3
			Echo-CoPilot	●	●	●	●	**54.0**

Conversely, the ReAct agent recovers stability (70%) by structuring evidence accumulation, while our proposed Echo-CoPilot with the SC mechanism further suppresses fluctuations (0.96 Avg Changes), achieving the highest overall stability (72%). This confirms that the SC mechanism effectively resolves borderline cases by grounding perspective-specific evidence in shared EchoKG logic.

Next, we conduct an error analysis by anatomy and views that exhibit the highest errors [16]. Table 3 (Left) demonstrates that our framework significantly mitigates the catastrophic failures of baselines in challenging regions. While the Qwen2-VL-7B-Biomed [16] suffers error rates of 83% and 71% on the Doppler views and PSAX great vessels, respectively, due to acoustic dropout, our tool-routed agent reduces these to 59.3% and 59.6%, proving highly robust against ambiguous acoustic windows.

Finally, the component ablations (Table 3, Right) show a steady improvement as components are added. The pure LLM baseline reaches 46.9% accuracy, adding tools and EchoKG (LLM+Tools) increases accuracy to 49.7%, and incorporating the ReAct loop further to derive an agentic framework (ReAct Agent) improves the performance to 52.3%. Finally, Echo-CoPilot with the SC mechanism gives the best result, which is a substantial improvement over the pure LLM baseline.

Limitations and the Challenge of Tool Variability. While our proposed Echo-CoPilot framework outperforms state-of-the-art baselines (Table 1), the overall performance ceiling reflects the severe vulnerability of current perceptual tools to domain shifts. We accept this modest accuracy margin as a necessary trade-off for explainable reasoning and verifiable guideline adherence. Notably, our 500-iteration stability test revealed that pure LLM logic errors are remarkably rare ($< 1\%$). Instead, *tool variability* remains the primary bottleneck: when foundation models extract noisy or conflicting perceptual data, the self-contrast mechanism exposes cross-perspective disagreements, leading to reasoning that is internally consistent but ultimately factually incorrect. Thus, while our architecture resolves the reasoning bottleneck, future work in agentic echocardiography must prioritize the deterministic reliability of underlying measurement tools.

5 Conclusion

We introduced Echo-CoPilot, an agentic framework that advances echocardiography interpretation from black-box prediction to guideline-constrained, verifiable reasoning. By enforcing clinical validity through EchoKG and resolving evidentiary conflicts via multi-perspective and self-contrast mechanisms, Echo-CoPilot achieves SOTA performance on MIMICEchoQA, significantly improving reliability near critical decision boundaries. While the scarcity of public video-based benchmarks currently limits evaluation to a single large-scale cohort, this work establishes a rigorous paradigm for auditable medical agents. Future efforts will focus on curating diverse, multi-center benchmarks to test generalization and advancing toward prospective clinical validation.

Acknowledgments. This work was supported by the Canadian Foundation for Innovation-John R. Evans Leaders Fund (CFI-JELF) program grant number 42816. Mitacs Accelerate program grant number AWD024298-IT33280. We also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2023-03575]. Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), [RGPIN-2023-03575].

Disclosure of Interests. The authors have no competing interests in the paper.

References

1. Agarwal, S., Ahmad, L., Ai, J., Altman, S., Applebaum, A., Arbus, E., Arora, R.K., Bai, Y., Baker, B., Bao, H., et al.: gpt-oss-120b & gpt-oss-20b model card. arXiv preprint arXiv:2508.10925 (2025)
2. Amadou, A.A., Zhang, Y., Piat, S., Klein, P., Schmuecking, I., Passerini, T., Sharma, P.: Echoapex: A general-purpose vision foundation model for echocardiography. arXiv preprint arXiv:2410.11092 (2024)
3. Bouchard, D., Chauhan, M.S.: Uncertainty quantification for language models: A suite of black-box, white-box, llm judge, and ensemble scorers. arXiv preprint arXiv:2504.19254 (2025)
4. Daghyani, M., Wang, L., Hashemi, N., Medhat, B., Abdelsamad, B., Velez, E.R., Li, X., Tsang, M.Y., Luong, C., Tsang, T.S., et al.: Echoagent: Guideline-centric reasoning agent for echocardiography measurement and interpretation. arXiv preprint arXiv:2511.13948 (2025)
5. Fallahpour, A., Ma, J., Munim, A., Lyu, H., Wang, B.: Medrax: Medical reasoning agent for chest x-ray (2025), https://arxiv.org/abs/2502.02673
6. Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., Vielhauer, J., Makowski, M., Braren, R., Kaissis, G., et al.: Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nature medicine **30**(9), 2613–2622 (2024)
7. Holste, G., Oikonomou, E.K., Tokodi, M., Kovács, A., Wang, Z., Khera, R.: Complete AI-enabled echocardiography interpretation with multitask deep learning. JAMA **334**(4), 306–318 (2025). https://doi.org/10.1001/jama.2025.8731

8. Huang, Y., Chen, Y., Xu, D., Zhao, C., Yue, W., Huang, Y.: Medreflect: Teaching medical llms to self-improve via reflective correction. arXiv preprint arXiv:2510.03687 (2025)
9. Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B.: Medsam2: Segment anything in 3d medical images and videos. arXiv preprint arXiv:2504.03600 (2025)
10. Mitchell, C., Rahko, P.S., Blauwet, L.A., Canaday, B., Finstuen, J.A., Foster, M.C., Horton, K., Ogunyankin, K.O., Palma, R.A., Velazquez, E.J.: Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the american society of echocardiography. Journal of the American Society of Echocardiography **32**(1), 1–64 (2019)
11. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., et al.: Video-based ai for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020)
12. Qin, Y., Gamage Nanayakkara, D.S., Li, X.: Multi-agent collaboration for integrating echocardiography expertise in multi-modal large language models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 358–368. Springer (2025)
13. Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J.: Ai in health and medicine. Nature medicine **28**(1), 31–38 (2022)
14. Sellergren, A., Kazemzadeh, S., Jaroensri, T., Kiraly, A., Traverse, M., Kohlberger, T., Xu, S., Jamil, F., Hughes, C., Lau, C., et al.: Medgemma technical report. arXiv preprint arXiv:2507.05201 (2025)
15. Shyr, C., Ren, B., Hsu, C.Y., Tinker, R.J., Cassini, T.A., Hamid, R., Wright, A., Bastarache, L., Peterson, J.F., Malin, B.A., et al.: A statistical framework for evaluating the repeatability and reproducibility of large language models. medRxiv pp. 2025–08 (2025)
16. Thapa, R., Li, A., Wu, Q., He, B., Sahashi, Y., Binder, C., Zhang, A., Athiwaratkun, B., Song, S.L., Ouyang, D., Zou, J.: How well can general vision-language models learn medicine by watching public educational videos? (2026), https://openreview.net/forum?id=u4PmZOmtko
17. Thapa, R., Li, A., Wu, Q., He, B., Sahashi, Y., Binder, C., Zhang, A., Athiwaratkun, B., Song, S.L., Ouyang, D., et al.: How well can general vision-language models learn medicine by watching public educational videos? arXiv preprint arXiv:2504.14391 (2025)
18. Ünlü, S., Duchenne, J., Mirea, O., Pagourelias, E.D., Bezy, S., Cvijic, M., Beela, A.S., Thomas, J.D., Badano, L.P., Voigt, J.U., et al.: Impact of apical foreshortening on deformation measurements: a report from the eacvi-ase strain standardization task force. European Heart Journal-Cardiovascular Imaging **21**(3), 337–343 (2020)
19. Valsangiacomo Büchel, E.R., Grosse-Wortmann, L., Fratz, S., Eichhorn, J., Sarikouch, S., Greil, G., Beerbaum, P., Bucciarelli-Ducci, C., Bonello, B., Sieverding, L., et al.: Indications for cardiovascular magnetic resonance in children with congenital and acquired heart disease: an expert consensus paper of the imaging working group of the aepc and the cardiovascular magnetic resonance section of the eacvi. European Heart Journal-Cardiovascular Imaging **16**(3), 281–297 (2015)
20. Vukadinovic, M., Tang, X., Yuan, N., Cheng, P., Li, D., Cheng, S., He, B., Ouyang, D.: Echoprime: a multi-video view-informed vision-language model for comprehensive echocardiography interpretation. arXiv preprint arXiv:2410.09704 (2024)

21. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., Cao, Y.: React: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629 (2022)
22. Zhang, W., Shen, Y., Wu, L., Peng, Q., Wang, J., Zhuang, Y., Lu, W.: Self-contrast: Better reflection through inconsistent solving perspectives. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics (2024)