Journal of Health Informatics

J Health Inform. 2026;18:e1577
DOI: https://doi.org/10.59681/2175-4411.v18.2026.1577

ARTIGO ORIGINAL

Concordância clínica de um chatbot clínico na interpretação de cineangiocoronariografia: análise comparativa com cardiologistas intervencionistas

Clinical agreement of a clinical chatbot in the interpretation of coronary angiography: a comparative analysis with interventional cardiologists

Concordancia clínica de un chatbot clínico en la interpretación de la angiografía coronaria: un análisis comparativo con cardiólogos intervencionistas

Josivan Soares Alves Júnior[1], Thayse Mota Alves[2], Debora Regina Alves Raposo[3], Alex Junior Vieira Sousa[1], Josué Luis Pereira Negreiros[1], Larissa Gomes Freire[1], Evely Laís Valença Melo[1], Cosme Michael Santos Farias[4] ✉

ORCID: 0000-0001-5294-3842 · 0000-0002-9227-138X · 0009-0006-6755-5918 · 0009-0001-7153-2077 · 0009-0009-7503-9296 · 0009-0007-3133-8003 · 0009-0007-0542-2367 · 0000-0002-1101-5764

[1] UNIFACISA – Centro Universitário, Campina Grande (PB), Brasil.
[2] Universidade de Pernambuco, Recife (PE), Brasil.
[3] Universidade Estadual da Paraíba, Campina Grande (PB), Brasil.
[4] Universidade Federal de Campina Grande, Campina Grande (PB), Brasil.

FICHA DO ARTIGO

TIPO
Artigo Original

IDENTIFICADOR
e1577

DOI
10.59681/2175-4411.v18.2026.1577

RECEBIDO
12/12/2025

APROVADO
02/03/2026

PUBLICADO
24/04/2026

PROCESSO DE AVALIAÇÃO
Double Blind Review

EDITOR RESPONSÁVEL
Juliano de Souza Gaspar

ACESSO
Aberto, CC BY 4.0

AUTOR CORRESPONDENTE
Cosme Michael Santos Farias
nutricosmemichael@gmail.com

LICENÇA

Este é um artigo publicado em acesso aberto sob a licença Creative Commons CC BY 4.0, que permite uso, distribuição e reprodução em qualquer meio, desde que o trabalho original seja corretamente citado.

CC BY 4.0

RESUMO

Introdução: Sistemas baseados em processamento de linguagem natural (PLN) têm sido propostos como suporte ao raciocínio clínico; contudo, seu uso em decisões terapêuticas complexas requer validação por especialistas. **Objetivo:** Avaliar a concordância clínica percebida por cardiologistas intervencionistas em relação às condutas sugeridas por um *chatbot* de IA a partir de achados de cineangiocoronariografia (cateterismo cardíaco, CATE). **Método:** Estudo comparativo, de abordagem mista, realizado no setor de hemodinâmica no estado da Paraíba. Quinze cardiologistas intervencionistas avaliaram três casos clínicos construídos a partir de exames de CATE de pacientes com síndrome coronariana aguda (SCA), atribuindo grau de concordância em escala *Likert* e justificando suas avaliações em respostas abertas analisadas por Discurso do Sujeito Coletivo (DSC). **Resultados:** Observou-se predomínio de concordância nas questões Q1 (92%) e Q3 (77%), indicando maior aceitabilidade das condutas em cenários mais lineares e alinhados a protocolos. Na Q2, 54% dos participantes discordaram da conduta, evidenciando maior divergência em decisões de revascularização mais complexas. **Conclusão:** As recomendações do *chatbot* apresentaram maior concordância percebida em cenários protocolizados, enquanto decisões que exigem maior individualização clínica concentraram discordâncias, reforçando a necessidade de uso supervisionado e orientado por julgamento especializado.

Descritores: Cardiologia; Saúde Cardiovascular; Tecnologia Biomédica.

ABSTRACT

Introduction: Natural language processing (NLP) systems have been proposed to support clinical reasoning; however, their use in complex therapeutic decisions requires expert validation. **Objective:** To assess interventional cardiologists' perceived clinical agreement with management recommendations generated by an AI chatbot based on coronary angiography findings. **Method:** A comparative mixed-methods study was conducted in the hemodynamics sector in the state of Paraíba, Brazil. Fifteen interventional cardiologists evaluated three clinical cases derived from coronary angiography examinations of patients with acute coronary syndrome (ACS), rating agreement on a Likert scale and providing open-ended justifications analyzed through the Collective Subject Discourse (CSD) technique. **Results:** Agreement predominated in Q1 (92%) and Q3 (77%), indicating higher acceptability in more linear, protocol-aligned scenarios. In Q2, 54% of participants disagreed, highlighting greater divergence in more complex revascularization decisions. **Conclusion:** The chatbot's recommendations showed higher perceived agreement in protocol-anchored scenarios, whereas decisions requiring greater clinical individualization concentrated disagreements, supporting supervised use under specialist judgment.

Keywords: Cardiology; Cardiovascular Health; Biomedical Technology.

RESUMEN

Introducción: Los sistemas basados en procesamiento de lenguaje natural (PLN) se han propuesto para apoyar el razonamiento clínico; sin embargo, su uso en decisiones terapéuticas complejas requiere validación por especialistas. **Objetivo:** Evaluar la concordancia clínica percibida por cardiólogos intervencionistas respecto a las conductas sugeridas por un chatbot de IA a partir de hallazgos de angiografía coronaria. **Método:** Se realizó un estudio comparativo de métodos mixtos en el sector de hemodinámica del estado de Paraíba, Brasil. Quince cardiólogos intervencionistas evaluaron tres casos clínicos derivados de exámenes de angiografía coronaria en pacientes con síndrome coronario agudo, calificando la concordancia en escala Likert y aportando justificaciones abiertas analizadas mediante Discurso del Sujeto Colectivo (DSC). **Resultados:** Se observó predominio de concordancia en Q1 (92%) y Q3 (77%), con mayor aceptabilidad en escenarios más lineales y alineados a protocolos. En Q2, 54% de los participantes discreparon, evidenciando mayor divergencia en decisiones de revascularización más complejas. **Conclusión:** Las recomendaciones del chatbot mostraron mayor concordancia percibida en escenarios protocolizados, mientras que decisiones que requieren mayor individualización clínica concentraron discrepancias, reforzando la necesidad de uso supervisado por especialistas.

Descriptores: Cardiología; Salud cardiovascular; Tecnología biomédica.

COMO CITAR ESTE ARTIGO: Alves Júnior JS, Alves TM, Raposo DRA, Sousa AJV, Negreiros JLP, Freire LG, et al. Concordância clínica de um chatbot clínico na interpretação de cineangiocoronariografia: análise comparativa com cardiologistas intervencionistas. J Health Inform 2026;18:e1577. doi: https://doi.org/10.59681/2175-4411.v18.2026.1577

1 INTRODUÇÃO

As doenças cardiovasculares (DCV) englobam patologias que acometem o coração e os vasos sanguíneos, caracterizando-se, em geral, por evolução crônica e, muitas vezes, assintomática nas fases iniciais. Para além dos fatores clínicos e biológicos, aspectos sociodemográficos e comportamentais estão fortemente associados à maior incidência de comorbidades cardiovasculares. Destacam-se, entre eles, ser do sexo masculino, autodeclarar-se de raça/cor branca, possuir nível médio de escolaridade e ter acesso a plano de saúde, bem como o uso abusivo de álcool e tabaco, a inatividade física e a adoção de uma alimentação inadequada, que se configuram como importantes determinantes para a progressão e o agravamento dessas condições [1,2].

Em decorrência desse conjunto de fatores, as DCV seguem como a principal causa de mortalidade no Brasil e no mundo, respondendo por aproximadamente 30% de todas as mortes globais [3]. No contexto brasileiro, dados da Sociedade Brasileira de Cardiologia (SBC) indicam que essas enfermidades foram responsáveis por cerca de 400 mil óbitos em 2022, muitos deles por causas potencialmente evitáveis, passíveis de detecção precoce por meio de exames clínicos, laboratoriais e de imagem. Após a pandemia de Covid-19, observou-se ainda uma redução expressiva na realização de procedimentos cardíacos, motivada pelo receio de contaminação em serviços de saúde, o que contribuiu para maior resistência da população em buscar diagnóstico e acompanhamento cardiológico [4, 5].

Diante desse cenário, o uso de inteligência artificial tem sido explorado como ferramenta de suporte ao raciocínio clínico, particularmente em aplicações baseadas em processamento de linguagem natural (PLN) e modelos de linguagem de grande escala, capazes de sintetizar informações e estruturar recomendações a partir de descrições textuais. No campo cardiovascular, revisões recentes descrevem aplicações em educação do paciente e, em menor proporção, em suporte à decisão clínica, mas ressaltam limitações como possibilidade de desinformação, sensibilidade ao contexto e necessidade de validação rigorosa antes de qualquer uso em cuidado individualizado [6,7].

Desse modo, a pesquisa busca responder à seguinte pergunta: um sistema de IA baseado em PLN, do tipo *chatbot,* gera recomendações clínicas que cardiologistas intervencionistas consideram concordantes e clinicamente adequadas quando confrontadas com casos construídos a partir de achados de cineangiocoronariografia? Para isso, as recomendações geradas pela IA serão avaliadas por um painel de especialistas em cardiologia intervencionista por meio de escala Likert e análise qualitativa das justificativas (DSC). Assim, o estudo tem como objetivo avaliar a concordância clínica percebida e a adequação clínica percebida das condutas sugeridas por um sistema de IA baseado em PLN, do tipo *chatbot*, a partir da análise de resultados de cineangiocoronariografia.

2 MÉTODO

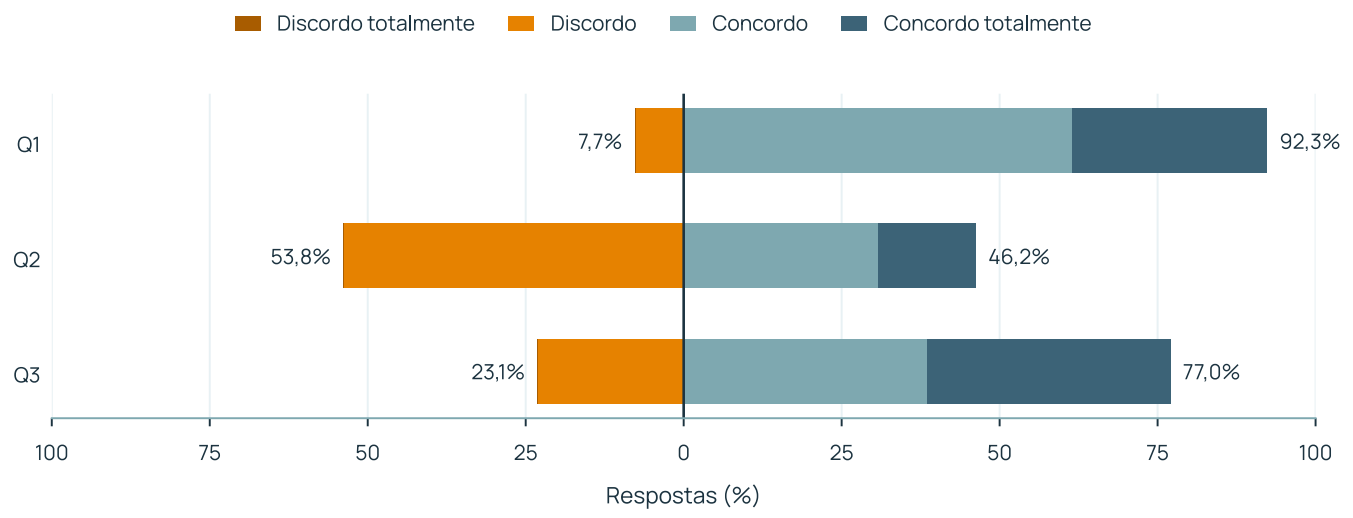
2.1 Delineamento do estudo

Trata-se de um estudo comparativo de método misto, com integração convergente, no qual dados quantitativos (escala *Likert*) e qualitativos (respostas abertas analisadas por Discurso do Sujeito Coletivo, DSC) foram coletados no mesmo período e integrados na interpretação para explicar convergências e divergências.

2.2 Local e período

A coleta de dados ocorreu em setembro de 2025, no setor de hemodinâmica de um hospital no estado da

Figura 1. Escala de concordância dos cardiologistas perante as condutas da IA.



Fonte: Elaborado pelos autores (2026).

https://doi.org/10.59681/2175-4411.v18.2026.1577

Paraíba, Brasil.

2.3 Participantes e critérios

Foram selecionados 15 cardiologistas intervencionistas, com experiência mínima de um ano em interpretação de cineangiocoronariografia (cateterismo cardíaco, CATE) e atuação em emergências cardiovasculares. A participação ocorreu mediante consentimento livre e esclarecido; a ausência de concordância formal impediu a participação por imperativo ético, e não como critério metodológico de exclusão.

2.4 Construção dos casos clínicos e elaboração do prompt

Foram construídos três casos clínicos a partir de exames de CATE de pacientes com diagnóstico de síndrome coronariana aguda (SCA) realizados no segundo semestre de 2024, com dados anonimizados e sem identificadores pessoais. Um *prompt* padronizado foi elaborado para solicitar ao *chatbot* a proposição de conduta clínica baseada nos achados apresentados, mantendo formato e nível de detalhamento consistentes entre os três casos.

2.5 Instrumento e procedimentos de coleta

As condutas produzidas pelo *chatbot* e os casos clínicos foram disponibilizados aos avaliadores por meio de formulário eletrônico em plataforma digital com controle de acesso e preservação do anonimato. Para cada caso, aplicou-se: (i) uma questão fechada em escala *Likert* ("Qual seu grau de concordância com a conduta proposta para este caso?"), com cinco opções (discordo totalmente, discordo, neutro, concordo, concordo totalmente); e (ii) uma questão aberta ("Qual sua opinião sobre a conduta gerada pelo *chatbot*?"). Ao final, foi incluída uma questão aberta exploratória: "Você consegue identificar quais condutas foram geradas por IA e quais por humanos?".

2.6 Análise dos dados

Os dados quantitativos foram analisados por estatística descritiva, com distribuição de frequências e representações gráficas. As respostas abertas foram analisadas por DSC, com identificação de expressões-chave, ideias centrais e construção de discursos-síntese representativos do pensamento coletivo. A integração dos achados quantitativos e qualitativos ocorreu na discussão, relacionando padrões de concordância/discordância às justificativas clínicas expressas nos discursos [8].

2.7 Aspectos éticos

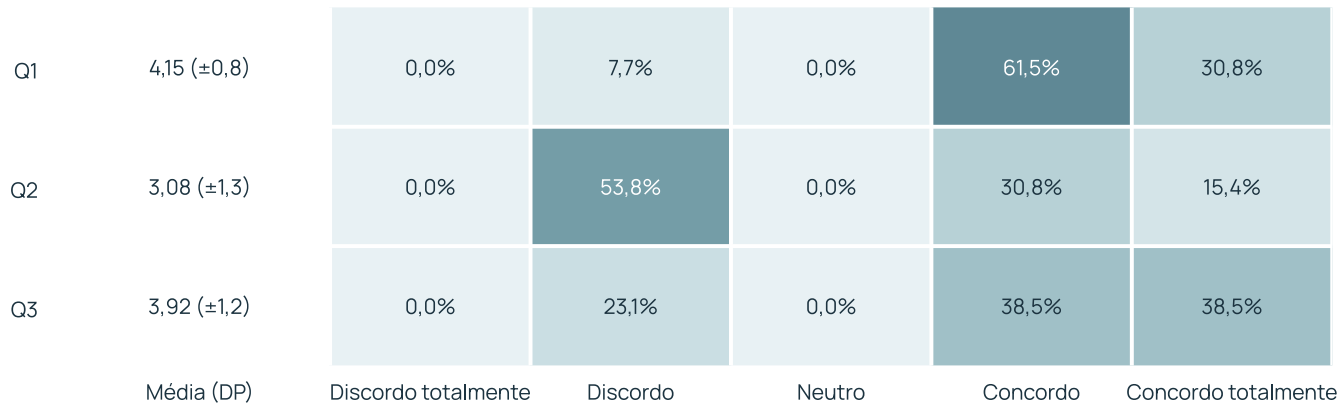
O projeto foi submetido ao Comitê de Ética em Pesquisa (CEP) do Centro de Ensino Superior e Desenvolvimento (CESED), com aprovação sob parecer nº 7.743.296 e CAAE: 86904925.3.0000.5175.

3 RESULTADOS/DISCUSSÃO

A Figura 1 ilustra a distribuição percentual das respostas dos avaliadores para três questões aplicadas aos casos clínicos gerados pelo *chatbot*, utilizando a Escala de *Likert*. Cada barra horizontal representa uma questão (Q1, Q2 e Q3), enquanto as cores indicam as categorias de resposta: "discordo totalmente", "discordo", "neutro", "concordo" e "concordo totalmente".

Na Questão 1 (Q1), observa-se uma predominância expressiva de avaliações positivas. A soma de "concordo" e "concordo totalmente" atinge 92%, distribuída visualmente em dois segmentos que ocupam quase toda a extensão positiva do gráfico. Há ainda 8% de respostas classificadas como "discordo", posicionadas próximas à área central, enquanto não há registros nas categorias de neutralidade, deixando a

Figura 2. Padrões de concordância e discordância: visualização métrica por mapa de calor.



Fonte: Elaborado pelos autores (2026).

porção completamente vazia para esta questão.

Em contrapartida, na Questão 2 (Q2), o padrão se inverte em relação à Q1. Aproximadamente 54% das respostas concentram-se nas categorias "discordo totalmente" e "discordo", formando uma extensa área na porção esquerda do gráfico, correspondente à região de respostas negativas. Já as categorias de concordância totalizam 46%, ocupando a parte direita da barra.

Já na Questão 3 (Q3), verifica-se novamente um predomínio de avaliações positivas, ainda que menos intenso que em Q1. As respostas "concordo" e "concordo totalmente" somam 77%, ocupando a maior parte da porção direita da barra. As respostas com discordância representam 23%, localizadas na região centro-esquerda do gráfico.

Após a análise da distribuição percentual das respostas apresentada no primeiro gráfico, que ilustra de forma direta a frequência das escolhas dos avaliadores nas cinco categorias da escala de *Likert*, torna-se necessário aprofundar essa visualização. Para isso, o segundo gráfico apresenta os mesmos dados organizados em um mapa de calor, permitindo observar com maior nitidez os padrões de intensidade das respostas e a relação entre as médias obtidas em cada questão e a concentração das frequências nas diferentes categorias.

A Figura 2 apresenta uma matriz de calor que representa a distribuição percentual das respostas dos avaliadores para três questões (Q1, Q2 e Q3), considerando as cinco categorias da Escala de *Likert:* "discordo totalmente", "discordo", "neutro", "concordo" e "concordo totalmente", sendo assim, cada célula contém o percentual correspondente à frequência da categoria e é preenchida com tonalidades variáveis de azul.

Em sequência, o gráfico apresenta à esquerda as médias e os desvios-padrão de cada questão, a fim de ter uma melhor análise desses dados. Com isso, a Q1 mostra uma média de 4,15 (±0,8), enquanto que a Q2 dispõe de uma média de 3,08 (±1,3) e, na Q3, pode ser visualizado o valor de 3,92 (±1,2), demonstrando uma variabilidade entre os níveis da escala.

Com base nisso, na Q1 não foi registrada resposta na categoria "discordo totalmente", ao mesmo tempo que 7,7% aparecem em "discordo". Entretanto, a maior concentração encontra-se nas categorias "concordo" e "concordo totalmente" que apresentam coloração intensa, com 61,5% e 30,8%, respectivamente.

Na Q2, observa-se a maior proporção de discordância entre as questões analisadas. Não há respostas em "discordo totalmente", mas 53,8% estão na categoria "discordo", representada por uma tonalidade mais escura devido à concentração elevada. Em contraste, não houve respostas neutras e as categorias de concordância registraram 30,8% em "concordo" e 15,4% em "concordo totalmente".

Por consonância, a Q3 mostra que não há registros de "discordo totalmente", mas 23,1% das respostas aparecem em "discordo". Ainda, como nas outras questões, não foram observadas respostas neutras; como contraponto, as categorias positivas mostraram uma divisão com 38,5% em "concordo" e 38,5% em "concordo totalmente", formando uma separação simétrica nessa categoria favorável.

Os resultados observados nas questões Q1 e Q3, que apresentaram predominância de avaliações positivas, sugerem maior concordância clínica percebida com as recomendações geradas pelo *chatbot* em cenários mais lineares e ancorados em condutas protocolizadas.

Quadro 1. 1º caso clínico – Qual sua opinião sobre a conduta gerada pelo chatbot?

Categoria	Ideia Central	Síntese da ideia predominante	Respondentes
A	Conduta correta, assertiva e baseada em evidências	Avaliação positiva da conduta, considerada clara, objetiva e adequada às evidências	R2, R5, R6, R9, R11
B	Necessidade de aprofundar investigação e condutas complementares	Reconhecimento da conduta inicial, mas com necessidade de maior investigação e ajustes terapêuticos	R1, R4, R7
C	Importância de condutas após a fase inicial	Ênfase na necessidade de planejamento após a abordagem inicial	R3

Fonte: Elaborado pelos autores (2026).

Quadro 2. 2º caso clínico – Qual sua opinião sobre a conduta gerada pelo chatbot?

Categoria	Ideia Central	Síntese da ideia predominante	Respondentes
A	Conduta assertiva e adequada	Concordância com a conduta baseada nos achados angiográficos	R9, R11
B	Individualização por idade, fragilidade e status clínico	Necessidade de adaptar a conduta conforme características do paciente	R1, R2, R8
C	Falta de contexto clínico e ajustes técnicos	Crítica à ausência de dados clínicos e inadequações técnicas	R4, R7
D	Protocolos e adaptação ao caso	Uso de diretrizes com necessidade de adequação à realidade clínica	R3

Fonte: Elaborado pelos autores (2026).

Evidências empíricas em condições cardíacas indicam que modelos conversacionais tendem a apresentar melhor desempenho em perguntas de baixa complexidade, enquanto, em cenários de maior complexidade clínica e necessidade de individualização, aumentam respostas incompletas, inconclusivas ou inadequadas quando comparadas à avaliação especializada. Por outro lado, a discrepância observada na Q2, marcada por maior índice de discordância, reforça que decisões de revascularização exigem integração de variáveis clínicas e anatômicas que frequentemente não estão integralmente representadas no enunciado, o que pode ampliar divergências entre recomendações automatizadas e julgamento especializado[9].

No entanto, a discrepância observada na Q2, marcada por maior índice de discordância entre os avaliadores, pode ser entendida à luz das limitações já descritas na literatura sobre o desempenho isolado da inteligência artificial em processos decisórios clínicos. A pesquisa afirma que, embora a IA represente uma ferramenta promissora, quase 30% dos estudos analisados apresentam sensibilidade inferior a 80%, evidenciando que os sistemas ainda não alcançam desempenho ideal quando operam sem o suporte do julgamento humano, reforçando que mesmo diante de situações clínicas mais complexas (como parece ter ocorrido no caso representado pela Q2) os modelos tendem a apresentar maior margem de erro ou interpretações insuficientes [10].

Além disso, o estudo de Roski et al. [11] afirma que a aplicação da IA em saúde deve priorizar o uso como inteligência assistida, valorizando a colaboração entre sistema automatizado e julgamento humano, especialmente diante da complexidade dos dados clínicos. Embora existam evidências de aplicações bem-sucedidas da IA em diagnóstico por imagem e predição, a pesquisa ressalta que a eficácia desses sistemas depende fortemente da qualidade, padronização e representatividade dos dados utilizados, o que limita seu desempenho quando essas condições não são garantidas.

Outro ponto apontado na literatura é a baixa transparência dos modelos de inteligência artificial que pode limitar sua adoção clínica, uma vez que respostas corretas nem sempre são acompanhadas de justificativas claras, o que compromete a confiança profissional. Além disso, esses sistemas podem reproduzir vieses oriundos dos dados de treinamento, sobretudo quando há limitações na qualidade ou representatividade das informações. No presente estudo, esses fatores ajudam a explicar possíveis divergências entre as condutas geradas pela IA e a avaliação dos cardiologistas, reforçando a importância da supervisão humana e da validação contínua [11].

Diante desses achados quantitativos, observa-se que, embora a IA tenha apresentado desempenho satisfatório em boa parte dos casos, persistem limitações perceptíveis que se refletem nas percepções dos especialistas. Assim, torna-se fundamental analisar, sob uma perspectiva qualitativa, os discursos dos cardiologistas, de modo a compreender de forma mais aprofundada as razões subjacentes às concordâncias e discordâncias observadas, o que será

Quadro 3. 3º caso clínico – Qual sua opinião sobre a conduta gerada pelo chatbot?

Categoria	Ideia Central	Síntese da ideia predominante	Respondentes
A	Anticoagulação e reavaliação	Indicação de suporte intensivo e reavaliação conforme contexto clínico	R1, R2
B	Clareza e completude da conduta	Ênfase na apresentação clara e completa da conduta clínica	R3, R11
C	Aprovação simples e direta	Aceitação direta da conduta apresentada	R4, R5
D	Protocolos e adaptação ao caso	Uso de protocolos com necessidade de individualização	R3

Fonte: Elaborado pelos autores (2026).

Quadro 4. Qual sua opinião sobre o uso da IA para essa finalidade?

Categoria	Ideia Central	Síntese da ideia predominante	Respondentes
A	Ferramenta de apoio, não substituto	IA como suporte ao raciocínio clínico, sem substituir o julgamento humano	R2, R7, R8, R9, R11
B	Limitações e necessidade de senso crítico	Restrição da IA em casos complexos e necessidade de discernimento	R1, R3, R4, R5
C	Utilidade prática	IA como ferramenta para esclarecimento de dúvidas	R6
D	Dificuldade de aplicabilidade	Incerteza quanto à utilidade prática da IA	R12

Fonte: Elaborado pelos autores (2026).

abordado na próxima seção por meio do DSC.

3.1 Síntese das percepções dos cardiologistas: Análise pelo DSC

Inicialmente, a análise qualitativa foi conduzida por meio da técnica do DSC. As respostas abertas dos cardiologistas, obtidas a partir das perguntas aplicadas aos casos clínicos e ao uso da inteligência artificial, foram analisadas individualmente, e delas foram extraídas as Expressões-Chave (ECH). Em seguida, essas expressões foram agrupadas em Ideias Centrais (IC), que deram origem às categorias analíticas, e, a partir da junção das falas semanticamente semelhantes, foram construídos os discursos, representando o pensamento coletivo dos participantes em relação a cada pergunta investigada.

A partir das categorias identificadas no Quadro 1, foram construídos os DSC, apresentados a seguir, os quais sintetizam o pensamento coletivo dos cardiologistas avaliadores em relação à questão analisada. Categorias A, B e C.

Categoria A: "Acredito que a conduta foi adequada, clara, objetiva e baseada em evidências. Foi assertiva e completa, com boas orientações e estratégias corretas de monitorização, diagnóstico e prevenção."

Categoria B: "Entendo que é necessário suporte intensivo e investigação diagnóstica aprofundada antes de propor intervenções como CDI. Também seria importante mencionar condutas adicionais, como o uso da amiodarona e, diante de achados específicos no cateterismo, considerar o controle com betabloqueador."

Categoria C: "Considero a conduta inicial clara, mas precisamos refletir sobre as estratégias que devem ser adotadas após essa primeira fase do atendimento."

No primeiro quadro, a maior parte dos respondentes concentrou-se na categoria "A – Conduta correta, assertiva e baseada em evidências", reconhecendo a proposta do *chatbot* como clara, objetiva e alinhada às boas práticas diagnósticas e de monitorização. As categorias "B – Necessidade de aprofundar investigação e condutas complementares" e "C – Importância de condutas após a fase inicial" indicam, porém, que os especialistas não enxergam a resposta da IA como solução completa, mas como ponto de partida que ainda exige investigação adicional, suporte intensivo e planejamento para fases posteriores do cuidado.

Esse padrão aparece de forma semelhante em estudos experimentais que avaliaram o desempenho do ChatGPT em cardiologia. Em um estudo de prova de conceito com condições cardíacas frequentes (AMSTELHEART-2), o modelo alcançou boa taxa de acerto em sintomas e condutas, especialmente em quadros mais lineares, mas os autores ressaltam a necessidade de revisão crítica por especialistas, justamente porque nuances clínicas importantes podem não ser captadas por um texto resumido [9]. De forma convergente, simulações com casos cardiológicos clínicos mostraram que o *chatbot* frequentemente oferece condutas aceitáveis, porém por vezes incompletas, o que obriga o clínico a complementar investigações, ajustar doses e definir estratégias de seguimento [12,13].

Em síntese, a discussão aponta que, em um primeiro caso em que a conduta parece mais protocolar, a IA produz um raciocínio considerado globalmente adequado, mas os cardiologistas mantêm uma postura de supervisão clínica rigorosa sobre o que foi gerado. Esse comportamento é coerente com a ideia de inteligência artificial como apoio ao raciocínio, e não como substituto do julgamento clínico, posição amplamente defendida na literatura contemporânea sobre IA em saúde [14,15].

A partir das categorias identificadas no Quadro 2, foram construídos os DSC, apresentados a seguir, os quais sintetizam o pensamento coletivo dos cardiologistas avaliadores em relação à questão analisada. Sendo assim, segue a ideia geral das quatro categorias A, B, C e D.

Categoria A: "Considero a conduta assertiva e de acordo com os achados do cateterismo, com manejo clínico acertado e orientações condizentes."

Categoria B: "A definição da conduta deve levar em conta o status clínico, fragilidade e idade do paciente. Em idosos frágeis, muitas vezes a melhor opção seria angioplastia ou tratamento clínico otimizado, enquanto em pacientes mais jovens ou com comorbidades como diabetes pode ser indicada a cirurgia de revascularização. Além disso, o cálculo do escore SYNTAX seria fundamental para auxiliar nessa decisão."

Categoria C: "Percebo que houve ausência de informações importantes no caso, como comorbidades e contexto do exame. Indicar cirurgia sem avaliar a possibilidade de angioplastia não é adequado. Além disso, não é recomendado prescrever inibidor de P2Y12 em pacientes com indicação cirúrgica, pelo risco de sangramento. Outras condutas, entretanto, foram apropriadas."

Categoria D: "A conduta segue protocolos institucionais e de associações, mas sempre precisa ser adaptada ao momento presente e à realidade clínica do paciente."

A distribuição das categorias revela um cenário de maior divergência entre os avaliadores. Embora parte dos profissionais classifique a resposta como "Conduta assertiva e adequada baseada nos achados angiográficos" (categoria A), outras categorias são francamente críticas. A categoria B destaca a necessidade de individualizar a conduta segundo idade, fragilidade e status clínico, a C aponta ausência de informações cruciais (comorbidades, contexto do exame) e questiona decisões como indicar cirurgia sem

considerar angioplastia ou prescrever inibidor de P2Y12 em paciente com indicação cirúrgica, a D, por sua vez, reconhece a importância dos protocolos, mas insiste na adaptação à realidade clínica do doente.

Esse padrão dialoga com a inversão observada na Q2 da escala de *Likert*, em que predominam respostas de discordância (54%), sugerindo que justamente esse caso que envolve decisão de revascularização mais complexa é o ponto em que a IA mais se distancia do consenso dos hemodinamicistas.

Resultados semelhantes aparecem em estudos que compararam recomendações de modelos de linguagem com decisões de *heart team* em doença coronariana. Em um estudo sobre decisão de revascularização, um LLM (*Large Language Model*) foi capaz de reproduzir parte das indicações do *heart team*, mas apresentou discordâncias relevantes em casos limítrofes, que exigiam ponderação de fragilidade, comorbidades e preferências do paciente [16]. Outro trabalho que avaliou o uso de grandes modelos de linguagem como suporte às decisões de revascularização mostrou boa concordância em cenários "livro-texto", porém desempenho inferior quando múltiplos fatores anatômicos e clínicos precisavam ser integrados à decisão cirúrgica ou percutânea, reforçando a necessidade de manter o *heart team* como instância central de decisão [17].

Além disso, uma simulação com *chatbot*, voltada para dúvidas em cardiologia clínica, mostrou que a ferramenta é capaz de oferecer respostas úteis, porém com erros técnicos pontuais que podem ter impacto em condutas sensíveis, reforçando a necessidade de validação por especialistas antes da aplicação à beira do leito [14]. Isso se aproxima da crítica dos participantes quanto à sugestão de antiagregante em candidato cirúrgico e à ausência de cálculos objetivos de risco para apoiar a decisão. Ao enfatizar fatores como idade avançada, fragilidade e diabetes, elementos que não estavam explicitamente presentes no *prompt*, os cardiologistas expõem uma limitação estrutural do modelo: ele raciocina sobre o que é descrito, não sobre aspectos que normalmente emergem na interação clínica com o paciente.

O Quadro 2 evidencia que, quando a decisão exige ponderar variáveis múltiplas e, em certa medida, subjetivas, o *chatbot* tende a reproduzir uma leitura mais protocolar do laudo angiográfico, enquanto o especialista reivindica espaço para julgamento individualizado. Esse descompasso é hoje um dos principais pontos de atenção na discussão sobre uso de IA em decisões terapêuticas complexas [16,17].

A partir das categorias identificadas no Quadro 3, foram construídos os DSC, apresentados a seguir, os quais sintetizam o pensamento coletivo dos cardiologistas avaliadores em relação à questão analisada. Categorias A, B, C e D.

Categoria A: "Indico suporte intensivo com anticoagulação plena, seguido de reavaliação com novo cateterismo. No entanto, é importante considerar o contexto clínico: se o paciente estiver assintomático, manteria a anticoagulação plena devido à carga trombótica em artéria coronária direita e indicaria programação cirúrgica, diante de um padrão de doença multiarterial grave."

Categoria B: "Considero que a conduta deve ser apresentada com clareza, contemplando os aspectos clínicos necessários. A interpretação exposta foi completa e adequada para o caso."

Categoria C: "Boa resposta, conduta aceitável e adequada ao contexto clínico apresentado."

Categoria D: "A conduta segue protocolos institucionais e de associações, mas sempre precisa ser adaptada ao momento presente e à realidade clínica do paciente."

No Quadro 3, observa-se novamente um predomínio de avaliações favoráveis em relação à conduta sugerida. As categorias "A – Anticoagulação e reavaliação" e "B – Clareza e completude da conduta" destacam que os avaliadores viram a proposta do *chatbot* como condizente com a necessidade de suporte intensivo, anticoagulação plena e reavaliação com novo cateterismo, principalmente diante de carga trombótica importante em artéria coronária direita e padrão de doença multiarterial grave. A categoria C expressa aceitação direta da conduta, e a D volta a reforçar a ideia de uso de protocolos associados à individualização caso a caso.

Esse resultado é coerente com o que se observa na Q3 da escala de *Likert*, na qual 77% dos participantes declararam concordar ou concordar totalmente com a conduta da IA, sugerindo que, nesse cenário específico, houve percepção de boa aderência às práticas de referência.

Estudos que testaram o ChatGPT em vinhetas cardiológicas reforçam essa interpretação. Em análise de casos clínicos em cardiologia, o modelo apresentou bom desempenho na definição de diagnóstico e manejo em cenário de síndrome coronariana, com especial acerto em descrever estratégias de anticoagulação e necessidade de reavaliação invasiva, embora ainda com inconsistências em detalhes de dose e escolha de fármacos em alguns casos [12]. De forma complementar, a avaliação de respostas do ChatGPT a questões sobre imagem cardíaca mostrou que o modelo é capaz de organizar condutas em linguagem clara e estruturada, o que sugere competência particular para sintetizar recomendações em formato compreensível e linear [15].

No presente estudo, a convergência entre a percepção de clareza, completude e a aceitação da conduta sugere que, quando o problema clínico é descrito com foco em um eixo central (trombo em artéria específica,

necessidade de reavaliação e programação cirúrgica), o *chatbot* consegue elaborar um plano reconhecido como coerente, que os especialistas se sentem à vontade para validar e, se necessário, refinar. A ênfase reiterada na necessidade de adaptar à realidade clínica indica, contudo, que essa aprovação se ancora na condição de a IA ser vista como suporte ao raciocínio, e nunca como decisão automática [12,15].

A partir das categorias identificadas no Quadro 4, foram construídos os DSC, apresentados a seguir, os quais sintetizam o pensamento coletivo dos cardiologistas avaliadores em relação à questão analisada. Sendo assim, segue a ideia geral das quatro categorias A, B, C e D.

Categoria A: "Considero a IA uma excelente ferramenta de apoio, útil para revisar protocolos, auxiliar no raciocínio clínico e ajudar a equipe a refletir sobre condutas. Entretanto, deve ser vista apenas como recurso complementar, não como única definidora das decisões, já que não substitui a análise à beira do leito e as nuances humanas que orientam a prática clínica."

Categoria B: "A IA apresenta limitações importantes, sobretudo em situações críticas que envolvem múltiplas variáveis complexas. Seu uso exige discernimento e conhecimento prévio, e há ferramentas médicas mais específicas e adequadas que o ChatGPT para essa finalidade. Além disso, é fundamental lembrar que tratamos o paciente e não apenas exames ou protocolos, sendo a avaliação clínica insubstituível."

Categoria C: "Vejo a IA como um recurso fundamental para tirar dúvidas e apoiar a prática, especialmente quando se busca rapidez e síntese de informações."

Categoria D: "Não compreendi bem qual seria a aplicabilidade dessa tecnologia no contexto proposto, o que gera dúvidas quanto ao seu real uso prático."

No eixo das percepções globais sobre a ferramenta, as categorias reunidas no Quadro 4 indicam que os cardiologistas enxergam a inteligência artificial como recurso adicional, e não como substituto do profissional. A IA é descrita como suporte para revisão de protocolos, organização do raciocínio clínico e esclarecimento de dúvidas pontuais, mas sua atuação é claramente condicionada à avaliação à beira do leito e à experiência do especialista. Ao mesmo tempo, os participantes destacam limitações importantes em cenários complexos, mencionam a existência de ferramentas médicas mais específicas do que um *chatbot* generalista e, em alguns casos, relatam dificuldade em visualizar a aplicabilidade prática da tecnologia no contexto da hemodinâmica.

Esse posicionamento encontra paralelo em estudos que exploraram o conhecimento, as atitudes e as práticas de profissionais de saúde frente a modelos de linguagem em cardiologia e em outras áreas. Pesquisas com equipes cardiovasculares apontam que os clínicos tendem a reconhecer o potencial da IA para apoiar a interpretação de exames, agilizar a busca de evidências e auxiliar na estratificação de risco, mas mantêm preocupações com acurácia, qualidade dos dados de entrada e necessidade de treinamento específico para uso seguro [18,19]. De modo semelhante, médicos de pronto atendimento pediátrico relataram ver ferramentas como o ChatGPT como promissoras para apoiar o raciocínio diagnóstico e a documentação clínica, ao mesmo tempo em que ressaltam riscos relacionados à segurança do paciente e à ausência de diretrizes claras para incorporação desses sistemas na rotina assistencial [20].

No contexto deste estudo, o conjunto dessas percepções sugere que os cardiologistas intervencionistas já incorporam uma leitura crítica e relativamente madura sobre o papel da IA: reconhecem a utilidade da ferramenta para organizar informações e revisar condutas, mas reafirmam que decisões em hemodinâmica permanecem dependentes do julgamento situacional, da integração multidisciplinar e da avaliação clínica direta do paciente. A tecnologia, portanto, é posicionada como apoio qualificado ao processo decisório, desde que utilizada sob supervisão e com clareza de seus limites [18-20].

4 CONCLUSÃO

Este estudo analisou o uso de um *chatbot* de inteligência artificial, baseado em linguagem natural, para sugerir condutas a partir do exame de cineangiocoronariografia de pacientes com síndrome coronariana aguda, confrontando essas recomendações com a opinião de cardiologistas intervencionistas de um serviço de hemodinâmica. De modo geral, a IA se aproximou das condutas adotadas na prática clínica quando os casos eram mais lineares e bem ancorados em protocolos, sendo percebida como um apoio útil ao raciocínio. Já nas situações em que a decisão exigia maior individualização, especialmente na escolha da estratégia de revascularização, a concordância diminuiu e ficaram mais evidentes as limitações do modelo para considerar, de forma integrada, o contexto clínico e as particularidades de cada paciente.

A partir dos discursos analisados, fica claro que os cardiologistas não rejeitam a IA, mas a enquadram no lugar que ela pode ocupar com segurança: o de ferramenta complementar. Eles reconhecem o potencial do *chatbot* para organizar informações, revisar condutas e estimular reflexão, ao mesmo tempo em que reafirmam que a avaliação à beira do leito, o julgamento especializado e a discussão em equipe continuam insubstituíveis. No âmbito da extensão universitária, essa experiência aproximou universidade e serviço, abriu espaço para debater o uso crítico e ético dessas tecnologias e contribuiu para a formação de estudantes e residentes mais atentos aos

riscos e às possibilidades da IA na cardiologia. Mesmo com as limitações de amostra e do uso de um modelo generalista, os achados sugerem que ferramentas baseadas em linguagem natural podem ser incorporadas de forma cuidadosa e supervisionada, como aliadas no processo formativo e assistencial, e não como substitutas da decisão clínica.

Contribuição dos autores. Os autores contribuíram igualmente em todas as etapas de construção do artigo: concepção e delineamento do estudo; coleta, análise e interpretação dos dados; redação e revisão crítica do manuscrito; e aprovação da versão final.

Financiamento. Os autores não declararam fonte de financiamento.

Disponibilidade de dados e materiais. Os autores não declararam a disponibilidade dos dados e materiais da pesquisa.

Conflitos de interesse. Os autores não declararam conflitos de interesse.

Declaração de uso de IA. Os autores não declararam uso de ferramentas de Inteligência Artificial.

REFERÊNCIAS

1. Brasil. Ministério da Saúde. Cerca de 400 mil pessoas morreram em 2022 no Brasil por problemas cardiovasculares. Brasília: Ministério da Saúde; 2023. [citado 2026 out 14]. Disponível em: https://bvsms.saude.gov.br/cerca-de-400-mil-pessoas-morreram-em-2022-no-brasil-por-problemas-cardiovasculares/
2. Gomes CS, Gonçalves RPF, Silva AG, Sá ACMGN, Alves FTA, Ribeiro ALP, et al. Factors associated with cardiovascular disease in the Brazilian adult population: National Health Survey, 2019. Rev Bras Epidemiol. 2021;24(Suppl 2):e210013. doi: https://doi.org/10.1590/1980-549720210013.supl.2
3. World Health Organization. Cardiovascular diseases (CVDs) [Internet]. Geneva: WHO; 2025. [citado 2026 out 14]. Disponível em: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
4. Agência Brasil. Doenças cardiovasculares matam 400 mil brasileiros por ano. Brasília: Agência Brasil; 2024. [citado 2026 out 14]. Disponível em: https://agenciabrasil.ebc.com.br/saude/noticia/2024-09/doencas-cardiovasculares-matam-400-mil-brasileiros-por-ano
5. Cerci RJ, Vitola JV, Paez D, Zuluaga A, Bittencourt MS, Sierra-Galan LM, et al. The Impact of COVID-19 on Diagnosis of Heart Disease in Latin America an INCAPS COVID Sub-analysis. Arq Bras Cardiol. 2022;118(4):745-53. doi: https://doi.org/10.36660/abc.20210388
6. Martins VBL, Thom LRF, Gonçalves JMS, Destefani AC. Avanços da inteligência artificial na cardiologia: uma revisão abrangente. Rev Ibero-Am Humanid Cienc Educ. 2024;10(4):2442-56. doi: https://doi.org/10.51891/rease.v10i4.13643
7. Ferreira Santos J, Ladeiras-Lopes R, Leite F, Dores H. Applications of large language models in cardiovascular disease: a systematic review. Eur Heart J Digit Health. 2025;6(4):540-53. doi: https://doi.org/10.1093/ehjdh/ztaf028
8. Genaro LE, Marconato JV, Pinotti FE, Valsecki Júnior A, Adas Saliba T, Lopez Rosell F. Discurso do sujeito coletivo: bases conceituais, enfoques e aplicações em pesquisas qualitativas. Rev Pesqui Qual. 2024;12(32):553-69. doi: https://doi.org/10.33361/rpq.2024.v.12.n.32.739
9. Harskamp RE, De Clercq L. Performance of ChatGPT as an AI-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). Acta Cardiol. 2024;79(3):358-66. doi: https://doi.org/10.1080/00015385.2024.2303528
10. Melo GBP, Araujo IBB, Guedes GP, Alves RN, Requeijo MJR. Um comparativo entre a inteligência artificial e os médicos radiologistas no campo da interpretação de exames de imagens. Res Soc Dev. 2024;13(11):e125131147411. doi: https://doi.org/10.33448/rsd-v13i11.47411
11. Roski J, Chapman W, Heffner J, Trivedi R, Del Fiol G, Kukafka R, et al. How artificial intelligence is changing health and health care. In: Whicher D, Ahmed M, Israni ST, et al., editors. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril [Internet]. Washington (DC): National Academies Press (US); 2023. [citado 2026 out 14]. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK605950/
12. Yavuz YE, Kahraman F. Evaluation of the prediagnosis and management of ChatGPT-4.0 in clinical cases in cardiology. Future Cardiol. 2024;20(4):197-207. doi: https://doi.org/10.1080/14796678.2024.2348898
13. Baladrón C, Sevilla T, Carrasco-Moraleja M, Gómez-Salvador I, Peral-Oliveira J, San Román JA. Assessing the accuracy of ChatGPT as a decision support tool in cardiology. Rev Esp Cardiol (Engl Ed). 2024;77(5):433-5. doi: https://doi.org/10.1016/j.rec.2023.11.011
14. Fernández-Cisnal A, Lopez-Ayala P, Miñana G, Boeddinghaus J, Mueller C, Sanchis J. Performance of an artificial intelligence chatbot with web search capability in cardiology-related assistance: a simulation study. Rev Esp Cardiol (Engl Ed). 2023;76(12):1065-7. doi: https://doi.org/10.1016/j.rec.2023.06.008
15. Monroe CL, Abdelhafez YG, Atsina K, Aman E, Nardo L, Madani MH. Evaluation of responses to cardiac imaging questions by the artificial intelligence large language model ChatGPT. Clin Imaging. 2024;112:110193. doi: https://doi.org/10.1016/j.clinimag.2024.110193
16. Mola S, Yıldırım A, Gül EB. Artificial Intelligence in Cardiac Treatment Decision-Making: An Evaluation of the Performance of ChatGPT Versus the Heart Team in Coronary Revascularization. Rev Cardiovasc Med. 2025;26(8):38705. doi: https://doi.org/10.31083/RCM38705

17. Sudri K, Motro-Feingold I, Ramon-Gonen R, Barda N, Klang E, Fefer P, et al. Enhancing Coronary Revascularization Decisions: The Promising Role of Large Language Models as a Decision-Support Tool for Multidisciplinary Heart Team. Circ Cardiovasc Interv. 2024;17(11):e014201. doi: https://doi.org/10.1161/CIRCINTERVENTIONS.124.014201

18. Jiang S, Ma L, Pan K, Zhang H. Knowledge, attitudes, and practices of cardiovascular health care personnel regarding coronary CTA and AI-assisted diagnosis: a cross-sectional study. J Glob Health. 2025;15:04103. doi: https://doi.org/10.7189/jogh.15.04103

19. Ozkan E, Tekin A, Ozkan MC, Cabrera D, Niven A, Dong Y. Global Health care Professionals' Perceptions of Large Language Model Use In Practice: Cross-Sectional Survey Study. JMIR Med Educ. 2025;11:e58801. doi: https://doi.org/10.2196/58801

20. AlGoraini Y, Alsayyali M, Alotaibi O, Almeshawi I, Alaifan F, Alrashed R. Perceptions of large language models in medical education and clinical practice among pediatric emergency physicians in Saudi Arabia: a multiregional cross-sectional study. Front Public Health. 2025;13:1634638. doi: https://doi.org/10.3389/fpubh.2025.1634638

NOTAS EDITORIAIS

O Journal of Health Informatics é uma publicação científica mantida pela Sociedade Brasileira de Informática em Saúde, associação científica sem fins lucrativos dedicada ao desenvolvimento da Informática em Saúde e da Saúde Digital no Brasil. O processo editorial é conduzido com independência científica, sem interferência institucional, comercial ou de patrocinadores, e segue as boas práticas internacionais de ética em publicação científica.

Este artigo foi avaliado por pares, em sistema duplo-cego, por especialistas da área. A equipe editorial usou ferramentas de inteligência artificial, com supervisão humana, na triagem inicial, na conferência do texto, das referências e dos números e, quando necessário, no redesenho de figuras, quadros e tabelas, aprovado pelos autores na prova. As ferramentas e as salvaguardas estão na Política de uso de inteligência artificial da revista. As decisões editoriais foram tomadas por pessoas.

SBIS Sociedade Brasileira de Informática em Saúde