Artigo Original
Deformidades
http://dx.doi.org/10.1590/S1808-185120252402293157

INTELIGÊNCIA ARTIFICIAL NO DIAGNÓSTICO DA ESCOLIOSE: UM ESTUDO COMPARATIVO ENTRE CHATGPT E CIRURGIÕES

ARTIFICIAL INTELLIGENCE IN SCOLIOSIS DIAGNOSIS: A COMPARATIVE STUDY BETWEEN CHATGPT AND SURGEONS

INTELIGENCIA ARTIFICIAL EN EL DIAGNÓSTICO DE LA ESCOLIOSIS: UN ESTUDIO COMPARATIVO ENTRE CHATGPT Y CIRUJANOS

Lucas Silveira Rabello de Oliveira[1], Rafael Carboni de Souza[1], André Evaristo Marcondes Cesar[1], Bruno Vieira Motter[1], Wilker Herkson de Almeida Oliveira[1], Guilherme Foizer[1], Gabrielle do Amaral Virginio Pereira[2], Luciano Miller Reis Rodrigues[1]

1. Faculdade de Medicina do ABC, Departamento de Ortopedia e Traumatologia, Cirurgia da Coluna, São Paulo, SP, Brasil.
2. Universidade de São Paulo (USP), Faculdade de medicina, São Paulo, SP, Brasil.

RESUMO

Objetivo: Este estudo explora a acurácia do ChatGPT na classificação e sugestão de condutas para escoliose idiopática do adolescente, avaliando o nível de concordância entre as respostas do modelo de inteligência artificial e as avaliações de especialistas em cirurgia de coluna vertebral. Procurando ajudar a responder a seguinte pergunta: É possível confiar no ChatGPT-4 (inteligência artificial de linguagem natural) para recomendar condutas diante de casos típicos do cotidiano, servindo como um auxílio para ortopedistas menos experientes ou até mesmo para médicos generalistas? A análise proposta busca identificar o potencial e as limitações da aplicabilidade da inteligência artificial no suporte ao diagnóstico e decisão clínica, sem treinamento prévio da plataforma. Métodos: Trata-se de um estudo transversal com cinco casos fictícios de escoliose idiopática apresentados ao ChatGPT, que forneceu a classificação de Lenke e uma sugestão de conduta para cada caso. Um painel de 37 cirurgiões avaliou as respostas, emitiu a melhor conduta e pontuou as recomendações do ChatGPT em uma escala Likert de 1 a 5, refletindo seu nível de concordância. Resultados: Em casos mais simples (Caso 1), o ChatGPT demonstrou alta concordância com os especialistas, com 97,3% dos cirurgiões concordando com a recomendação de "cirurgia de instrumentação" (AC1=0,95). Entretanto, em casos mais complexos (Casos 3 e 5), a concordância foi significativamente menor, com apenas 11,1% e 18,8% dos especialistas aceitando as recomendações da IA, respectivamente. A precisão do modelo na classificação de Lenke foi consistente em todos os casos, evidenciando sua capacidade de aplicar critérios padronizados. Não houve correlação significativa entre a experiência dos cirurgiões e o nível de concordância com o software. Conclusão: O ChatGPT demonstrou potencial como ferramenta auxiliar no diagnóstico e planejamento terapêutico de escoliose, especialmente na classificação, porém ainda não está pronta para ser utilizada de maneira confiável e replicável, especialmente em casos mais complexos, especialmente em considerar nuances clínicas e fatores individuais do paciente. Apesar de promissora, a adoção dessa tecnologia poderá complementar o julgamento clínico, mas ainda requer supervisão e não substitui o papel da avaliação médica especializada no cenário atual. ***Nível de Evidência IV; Estudos Observacionais Descritivos.***

Descritores: Inteligência Artificial; Escoliose; Coluna Vertebral; Fusão vertebral; Curvaturas da Coluna Vertebral; Estudo Comparativo.

ABSTRACT

Objective: This study explores the accuracy of ChatGPT in classifying and suggesting approaches for adolescent idiopathic scoliosis, assessing the level of agreement between the artificial intelligence model's responses and the evaluations of spine surgery specialists. It aims to help answer the following question: Is it possible to trust ChatGPT-4 (natural language artificial intelligence) to recommend approaches for typical everyday cases, aiding less experienced orthopedists or even general practitioners? The proposed analysis seeks to identify the potential and limitations of applying artificial intelligence to support diagnosis and clinical decision-making without prior training of the platform. Methods: This is a cross-sectional study involving five fictitious cases of idiopathic scoliosis presented to ChatGPT, which provided the Lenke classification and a suggested approach for each case. A panel of 37 surgeons evaluated the responses, determined the best approach, and scored ChatGPT's recommendations on a Likert scale from 1 to 5, reflecting their level of agreement. Results: In simpler cases (Case 1), ChatGPT showed high agreement with the specialists, with 97.3% of the surgeons agreeing with the recommendation of "instrumentation surgery" (AC1=0.95). However, agreement was significantly lower in more complex cases (Cases 3 and 5), with only 11.1% and 18.8% of the specialists accepting the AI's recommendations, respectively. The model's accuracy in the Lenke classification was consistent across all cases, demonstrating its ability to apply standardized criteria. There was no significant correlation between the surgeons' experience and their level of agreement with the software. Conclusion: ChatGPT showed potential as an auxiliary tool in the diagnosis and therapeutic planning of scoliosis, particularly in classification, but it is not yet ready to be used reliably and consistently, especially in more complex cases, particularly when considering clinical nuances and individual patient factors. Although promising, the adoption of this technology can complement clinical judgment but still requires supervision and does not replace the role of specialized medical evaluation in the current scenario. ***Level of Evidence IV; Descriptive Observational Studies.***

Keywords: *Artificial Intelligence; Scoliosis; Spine; Spinal Fusion; Spinal Curvatures; Comparative Study.*

Estudo realizado no Hospital Estadual Mário Covas, R. Dr. Henrique Calderazzo, 321, Santo André, São Paulo, SP, Brasil. 09190-615
Correspondence: Lucas Silveira Rabello de Oliveira. R. Dr. Henrique Calderazzo, 321, Santo André, São Paulo, SP, Brasil. 09190-615 doutorrabello@gmail.com

CC BY

Recebido em 30/01/2025 aceito em 12/03/2025
Revisado por: Marcelo Wajchenberg

RESUMEN

Objetivo: Este estudio explora la precisión de ChatGPT en la clasificación y sugerencia de conductas para la escoliosis idiopática del adolescente, evaluando el nivel de concordancia entre las respuestas del modelo de inteligencia artificial y las evaluaciones de especialistas en cirugía de columna vertebral. Busca ayudar a responder la siguiente pregunta: ¿Es posible confiar en ChatGPT-4 (inteligencia artificial de lenguaje natural) para recomendar conductas en casos típicos del día a día, sirviendo como una ayuda para ortopedistas menos experimentados o incluso médicos generalistas? El análisis propuesto busca identificar el potencial y las limitaciones de la aplicabilidad de la inteligencia artificial en el apoyo al diagnóstico y la toma de decisiones clínicas, sin entrenamiento previo de la plataforma. Métodos: Se trata de un estudio transversal con cinco casos ficticios de escoliosis idiopática presentados a ChatGPT, que proporcionó la clasificación de Lenke y una sugerencia de conducta para cada caso. Un panel de 37 cirujanos evaluó las respuestas, emitió la mejor conducta y puntuó las recomendaciones de ChatGPT en una escala Likert de 1 a 5, reflejando su nivel de concordancia. Resultados: En casos más simples (Caso 1), ChatGPT demostró una alta concordancia con los especialistas, con el 97,3% de los cirujanos de acuerdo con la recomendación de "cirugía de instrumentación" (AC1=0,95). Sin embargo, en casos más complejos (Casos 3 y 5), la concordancia fue significativamente menor, con solo el 11,1% y el 18,8% de los especialistas aceptando las recomendaciones de la IA, respectivamente. La precisión del modelo en la clasificación de Lenke fue consistente en todos los casos, evidenciando su capacidad para aplicar criterios estandarizados. No hubo una correlación significativa entre la experiencia de los cirujanos y el nivel de concordancia con el software. Conclusión: ChatGPT demostró potencial como herramienta auxiliar en el diagnóstico y la planificación terapéutica de la escoliosis, especialmente en la clasificación, pero aún no está listo para ser utilizado de manera confiable y replicable, particularmente en casos más complejos, especialmente al considerar matices clínicos y factores individuales del paciente. Aunque prometedora, la adopción de esta tecnología puede complementar el juicio clínico, pero aún requiere supervisión y no reemplaza el papel de la evaluación médica especializada en el escenario actual. ***Nivel de Evidencia IV; Estudios Observacionales Descriptivos.***

Descriptores: *Inteligencia Artificial; Escoliosis; Columna Vertebral; Fusión Vertebral; Curvaturas de la Columna Vertebral; Estudio Comparativo.*

INTRODUÇÃO

A Inteligência Artificial (IA) tem atraído atenção crescente na área médica, prometendo transformações radicais na maneira como os cuidados de saúde são entregues. No entanto, a implementação bem-sucedida da IA na prática clínica requer uma avaliação rigorosa de sua eficácia e confiabilidade.[1]

O termo "machine learning" foi cunhado por Arthur Samuel em 1959,[2] definindo um campo da inteligência artificial (IA) no qual computadores aprendem automaticamente a partir da acumulação de dados. Essa técnica tem sido amplamente aplicada na análise de grandes volumes de dados, particularmente para o processamento de imagens.[3] Ao contrário de softwares tradicionais, que necessitam de instruções específicas para realizar uma tarefa, o deep learning permite que os sistemas reconheçam padrões de maneira autônoma e realizem previsões.[4]

De acordo com Pesapane et al., a aplicação da IA transformará metodologias de trabalho em várias profissões, incluindo a medicina, sendo que na radiologia essa mudança ocorrerá de maneira mais acelerada do que em outras especialidades médicas.[5]

O ChatGPT é uma inteligência artificial de linguagem natural bastante avançada, com o objetivo de se tornar uma *large language model* (LLM), desenvolvido pela OpenAI, sendo a mais utilizada atualmente e talvez, na sua versão atual (ChatGPT-4o), seja a mais completa e precisa neste modelo de proposta.[6]

O modelo aprendeu com uma ampla gama de textos, como livros, artigos, websites e outros materiais educativos e científicos. No entanto, é importante mencionar que, embora o modelo possa gerar respostas informadas baseadas na informação que aprendeu, não tem acesso a dados médicos individuais, confidenciais ou atualizados. As IA's generativas geram respostas com base nas relações probabilísticas entre as palavras e frases em sua base de treinamento.[6,7]

O ChatGPT não tem conhecimento ou consciência por si mesmo, mas usa padrões aprendidos para produzir respostas relevantes e coerentes, também é preciso ressaltar que o software tem limitações quanto a análise crítica das informações apresentadas.[1,6] Portanto, devido ao fato das condutas médicas serem extremamente individualizadas, dependendo de vários fatores intrínsecos ao paciente, além da experiência adquirida na trajetória de cada cirurgião, consideramos pouco provável que o ChatGPT forneça informações confiáveis de maneira que seja possível adiar ou até mesmo dispensar uma avaliação médica especializada.

Com o advento da inteligência artificial e sua crescente aplicação em diversos campos, o impacto da tecnologia AI, particularmente o modelo de linguagem ChatGPT, na prática médica e cirúrgica se tornou um tema de investigação crucial. Especificamente, a aplicação desta tecnologia avançada na cirurgia da coluna vertebral é o foco deste estudo. Além disso, He et al. observaram que o ChatGPT pode melhorar o suporte intraoperatório, fornecendo informações de navegação cirúrgica em tempo real e monitoramento de parâmetros fisiológicos, além de auxiliar na orientação de reabilitação pós-operatória, otimizando a coleta e análise de dados do paciente, melhorar a comunicação entre cirurgiões da coluna vertebral, pacientes e seus familiares, e contribuir para o processo de planejamento cirúrgico.[8]

O estudo conduzido explora a acurácia na classificação e sugestão de condutas para escoliose idiopática do adolescente, avaliando o nível de concordância entre as respostas do *chatbot* e as avaliações de especialistas em cirurgia de coluna vertebral. Procurando ajudar a responder a seguinte pergunta: É possível confiar no ChatGPT-4 (inteligência artificial de linguagem natural) para recomendar condutas diante de casos típicos do cotidiano, servindo como um auxílio para ortopedistas menos experientes ou até mesmo para médicos generalistas? A análise proposta busca identificar o potencial e as limitações do uso da inteligência artificial na prática clínica, contribuindo para a literatura existente. Permitindo avaliar a aplicabilidade da inteligência artificial no suporte ao diagnóstico e à decisão clínica no cenário atual, sem treinamento prévio da plataforma.

MÉTODO

O estudo conduzido utiliza uma investigação transversal usando uma metodologia quantitativa para avaliar a concordância entre o ChatGPT-4o e os especialistas em cirurgia da coluna vertebral em cinco casos fictícios de Escoliose Idiopática do Adolescente (EIA). A análise focou na comparação das opções de tratamento sugeridas pelos profissionais e pela Inteligência Artificial (IA), bem como no nível de concordância dos cirurgiões com as explicações fornecidas pela IA. Os casos apresentados foram analisados pela IA e fornecidos uma classificação de Lenke e uma sugestão de conduta para cada caso. Um painel de especialistas foi convidado a avaliar cada resposta do ChatGPT-4 utilizando a escala Likert de 5 pontos, que varia de "discordo fortemente" a "concordo fortemente".

Coleta de Dados

A coleta de dados foi efetuada por meio de um questionário online, contendo os casos clínicos a serem analisados e as respostas do ChatGPT4, quanto **às classificações** da patologia e à conduta médica sugerida. As respostas foram analisadas de forma anônima e divulgadas e com dados agrupados, sem identificação dos participantes. Além disso, por ser um estudo sem intervenção clínica e metodologia observacional, baseado na segurança das informações e sem identificação dos pacientes, não haverá risco adicional promovido por essa pesquisa. Todos os pesquisadores envolvidos nesse trabalho se comprometeram individualmente e coletivamente a utilizar os dados dessa pesquisa apenas com fins descritivos, com respeito ao sigilo e confidencialidade dos dados coletados.

Participaram do estudo 37 cirurgiões, incluindo ortopedistas especialistas em cirurgia da coluna vertebral, fellows de cirurgia da coluna vertebral, residentes de neurocirurgia e neurocirurgiões, recrutados de várias instituições de saúde. Cada participante recebeu cinco casos clínicos detalhados de EIA, incluindo informações necessárias para a classificação de Lenke e para a decisão terapêutica.

Todos os participantes assinaram e concordaram com o TCLE, de acordo com Resolução CNS N° 466 de 2012. O presente estudo foi submetido à Plataforma Brasil e avaliado pelo Comitê de Ética em Pesquisa, recebendo o Número do Parecer: 7.070.552 e o CAAE: 81977424.5.0000.0082. Conforme o disposto na Resolução CNS nº 510 de 07 de Abril de 2016, parágrafo único, o projeto foi dispensado de análise ética, não sendo necessária avaliação adicional por parte do comitê.

Após a apresentação de cada caso, os cirurgiões foram solicitados a escolher a melhor opção de tratamento dentre as seguintes alternativas: a) colete TLSO; b) colete CTLSO; c) cirurgia de instrumentação; d) expectante - com acompanhamento seriado; e) expectante – alta.

Em seguida, os participantes leram a resposta do ChatGPT, que incluía a classificação de Lenke e a proposta de tratamento. Por fim, os cirurgiões avaliaram seu nível de concordância com a resposta da IA em uma escala Likert de 1 a 5, onde: 1 = discordo fortemente; 2 = discordo; 3 = concordo em parte; 4 = concordo; 5 = concordo fortemente.

Análise de Dados

As análises foram realizadas utilizando os softwares estatísticos Jamovi[9,10] e Excel. As variáveis categóricas são apresentadas como frequência e porcentagens (%) e as variáveis contínuas como média e desvio padrão (M±DP). Valores de $p<0,05$ são considerados estatisticamente significantes.

Para a análise dos dados, calculamos as frequências absolutas e relativas das opções de tratamento escolhidas pelos cirurgiões em cada caso, comparando-as com a opção sugerida pelo ChatGPT. As respostas "colete TLSO" e "colete CTLSO" foram agrupadas como "tratamento com colete", pois ambas representam modalidades semelhantes de tratamento ortótico.

Devido à natureza nominal dos dados e ao número de avaliadores, utilizamos o Coeficiente AC1 de Gwet para avaliar a concordância entre os cirurgiões e a IA. Este coeficiente é adequado para dados categóricos com múltiplos avaliadores e é menos suscetível ao paradoxo da prevalência que pode afetar outros coeficientes, como o Kappa de Fleiss. O Coeficiente AC1 varia de 0 a 1, onde valores próximos a 1 indicam alta concordância além do acaso, e valores próximos a 0 indicam concordância equivalente ao acaso. Sua interpretação é semelhante à do coeficiente Kappa, com valores entre 0,81 e 1,00 indicando concordância quase perfeita e entre 0,61 e 0,80 indicando concordância substancial.

Para avaliar a consistência nas decisões clínicas dos cirurgiões, calculamos a proporção de concordância observada (P_o) em cada caso. Está métrica reflete o grau de consenso clínico, indicando a frequência com que os cirurgiões escolheram a mesma opção de tratamento para um caso específico. Uma alta P_o indicou um consenso clínico mais estabelecido, enquanto uma P_o mais baixa apontou maior variabilidade nas preferências de tratamento. Esta análise permitiu identificar padrões de concordância entre os especialistas e comparar a variabilidade de acordo com a complexidade de cada caso clínico.

Os cirurgiões avaliaram as respostas do ChatGPT em uma escala Likert de 1 a 5 para indicar seu nível de concordância com as recomendações da IA. Para cada caso, foram calculadas medidas de tendência central (média, mediana, moda) e dispersão (desvio padrão e amplitude interquartil) para facilitar a visualização dos padrões de concordância e identificar potenciais polarizações nas respostas.

Investigamos se a experiência dos cirurgiões influenciava o nível de concordância com o ChatGPT, calculando a correlação de Spearman entre os anos de experiência e o nível de concordância em cada caso. A correlação de Spearman, adequada para dados não paramétricos, avalia a força e a direção da associação entre duas variáveis ordinais ou intervalares.

Identificamos casos em que os cirurgiões escolheram a mesma opção de tratamento que a IA, mas apresentaram baixo nível de concordância, e elaboramos hipóteses para explicar essas discrepâncias. Por fim, para avaliar se a formação profissional (ortopedista, fellow ou neurocirurgião) influenciou o nível de concordância, utilizamos o teste de Kruskal-Wallis.

RESULTADOS

Análise dos dados gerais

A Tabela 1 apresenta os anos de experiência dos 37 profissionais participantes do estudo. O grupo foi composto por 9 profissionais (24,3%) com menos de dois anos de experiência em cirurgia de coluna, 9 profissionais (24,3%) com de dois a cinco anos de experiência, 10 profissionais (27%) com menos de cinco a dez anos de experiência e 9 profissionais (24,3%) com mais de dez anos de experiência. A distribuição homogênea dos participantes não apresentou diferença estatisticamente significativa entre as proporções de profissionais nos diferentes grupos de experiência ($\chi^2 = 0,0811$, $p = 0,994$), indicando que a amostra é representativa e equilibrada.

Caso 1

No Caso 1, o ChatGPT recomendou "cirurgia de instrumentação" como a melhor opção de tratamento. Dos 37 cirurgiões, 36 (97,3%) também selecionaram "cirurgia de instrumentação", enquanto apenas um (2,7%) escolheu "colete TLSO". O Coeficiente AC1 de Gwet entre os cirurgiões e o ChatGPT foi de 0,95, indicando concordância quase perfeita. A proporção de concordância direta foi de 97,3%, refletindo um altíssimo nível de concordância entre os cirurgiões e o ChatGPT. Figura 1

A proporção de concordância entre os cirurgiões observada (P_o) entre os cirurgiões neste caso foi de 0,946, indicando alta concordância. Isso significa que 94,6% das possíveis combinações de respostas entre os cirurgiões foram concordantes. Essa alta concordância reflete um consenso clínico bem estabelecido sobre o tratamento adequado para o Caso 1.

A média do nível de concordância na escala Likert foi de 4,27 (desvio padrão = 0,769150), com mediana de 4,0 e moda de 5. A Amplitude Interquartil (IQR) foi de 1,0 (Q1 = 4,0; Q3 = 5,0), com variância: 0,59. Nenhum cirurgião atribuiu notas nos níveis 1 ou 2, reforçando a aceitação geral da recomendação do ChatGPT. O gráfico de barras (Figura 2) demonstra que a maioria dos cirurgiões atribuiu notas altas de concordância (4 e 5), reforçando essa tendência.

Observou-se uma correlação negativa significativa entre os anos de experiência e o nível de concordância (rho = -0,29; p = 0,040), sugerindo que cirurgiões com menos experiência tendem a concordar mais com o ChatGPT.

Caso 2

No Caso 2, o ChatGPT recomendou "colete TLSO" como a melhor opção de tratamento. Ao agrupar "colete TLSO" e

Tabela 1. Participantes do estudo (N=37).

Anos de experiência em cirurgia de coluna	n	%
Menos de 2 anos	9	24,3%
Entre 2 a 5 anos	9	24,3%
Entre 5 a 10 anos	10	27,0%
Mais de 10 anos	9	24,3%

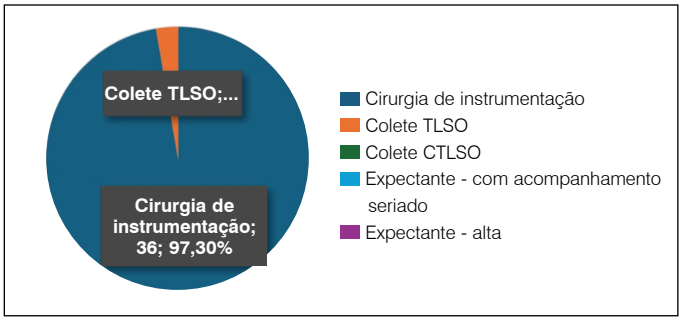

Figura 1. Proporção das respostas dos participantes para o Caso 1.

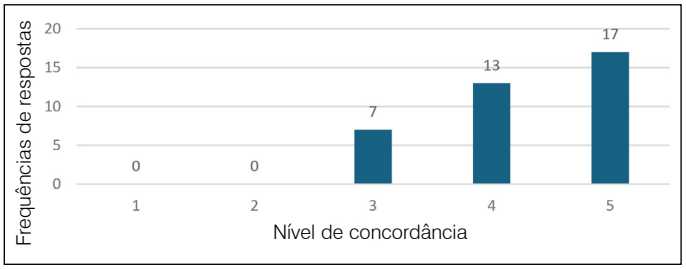

Figura 2. Proporção de respostas da escala Likert para o caso 1

"colete CTLSO" como "tratamento com colete", 33 dos 36 cirurgiões (91,7%) concordaram com a recomendação da IA. Os outros 3 cirurgiões (8,3%) escolheram opções diferentes, como "cirurgia de instrumentação" ou "expectante com acompanhamento seriado". O coeficiente AC1 de Gwet entre os cirurgiões e o ChatGPT foi de 0,86, indicando concordância quase perfeita. A proporção de concordância direta foi de 91,7%. (Figura 3).

A proporção de concordância entre cirurgiões observada (P_o) entre os cirurgiões neste caso foi de 0,722, indicando concordância substancial. Isso significa que 72,2% das combinações de respostas entre os cirurgiões foram concordantes.

A média do nível de concordância na escala Likert foi de 2,94 (desvio padrão = 1,433167), com mediana de 3,5 e moda de 4,0. A Amplitude Interquartil (IQR) foi de 2,25 (Q1 = 1,75; Q3 = 4,0), com variância: 2,05, indicando que a resposta mais frequente foi "concordo parcialmente", mostrando que as opiniões dos cirurgiões variaram, embora haja uma tendência à concordância. (Figura 4).

Não houve correlação significativa entre os anos de experiência dos cirurgiões e o nível de concordância com o ChatGPT (ρ = -0,09; p = 0,58).

Caso 3

No Caso 3, o ChatGPT sugeriu "colete TLSO" como tratamento. Dos 36 cirurgiões que responderam a este caso, 32 (88,9%) escolheram "cirurgia de instrumentação", enquanto 4 (11,1%) optaram por "tratamento com colete" e nenhum escolheu "expectante". O coeficiente AC1 de Gwet entre os cirurgiões e o ChatGPT foi de 0,03, indicando discordância. (Figura 5).

A proporção de concordância direta entre os cirurgiões foi de 88,9%, refletindo uma alta concordância na escolha de "cirurgia de instrumentação". A proporção de concordância observada (P_o) entre os cirurgiões no Caso 3 foi de aproximadamente 0,792, indicando concordância substancial a quase perfeita. Isso significa que 79,2% das combinações possíveis de respostas entre os cirurgiões foram concordantes, refletindo a alta concordância na escolha de "cirurgia de instrumentação" como tratamento adequado.

A média do nível de concordância na escala Likert foi de 1,89 (desvio padrão = 0,863174), com mediana de 2,0 e moda de 2. A Amplitude Interquartil (IQR) foi de 1,0 (Q1 = 1,0; Q3 = 2,0), com variância: 0,74, com as respostas mais frequentes tendendo a discordância. A maioria dos cirurgiões atribuiram notas baixas (1 ou 2), refletindo a discordância geral com a recomendação da IA (Figura 5). Não houve correlação significativa entre os anos de experiência e o nível de concordância (ρ = -0,05; p = 0,71). (Figura 6).

Caso 4

As escolhas dos cirurgiões foram: "cirurgia de instrumentação" por 10 cirurgiões (30,3%), "colete TLSO" por 1 cirurgião (3,0%), "colete CTLSO" por 1 cirurgião (3,0%), "expectante - acompanhamento seriado" por 21 cirurgiões (63,6%) e "expectante - alta" por 4 cirurgiões (10,8%). Portanto, 10 cirurgiões (30,3%) concordaram com a recomendação da IA. (Figura 7).

A proporção de concordância observada (P_o) entre os cirurgiões foi de 0,393, indicando concordância moderada. Isso significa que 39,3% das combinações de respostas entre os cirurgiões foram concordantes.

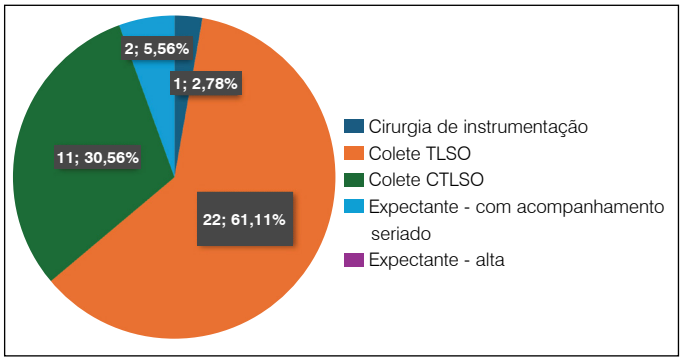

Figura 3. Proporção das respostas dos participantes para o Caso 2.

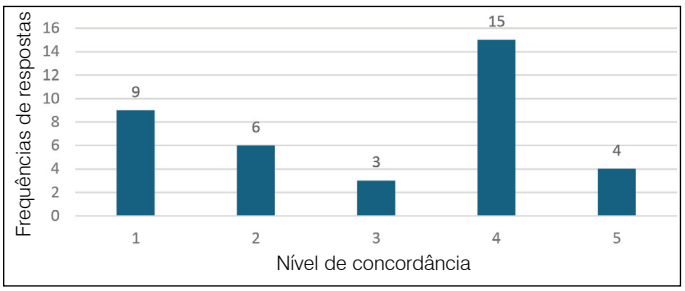

Figura 4. Proporção de respostas da escala Likert para o caso 2.

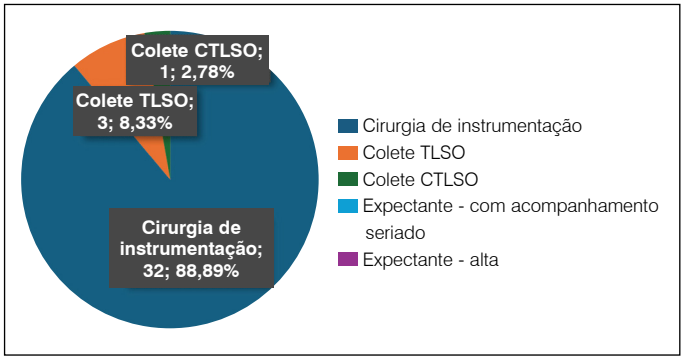

Figura 5. Proporção das respostas dos participantes para o Caso 3.

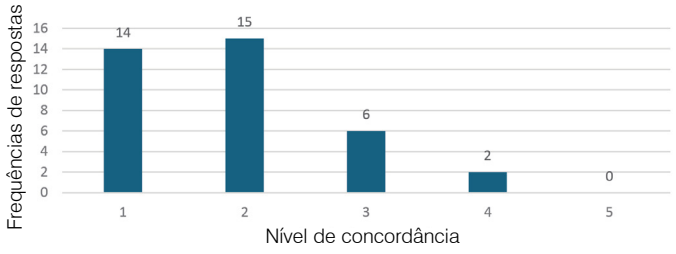

Figura 6. Proporção de respostas da escala Likert para o caso 3.

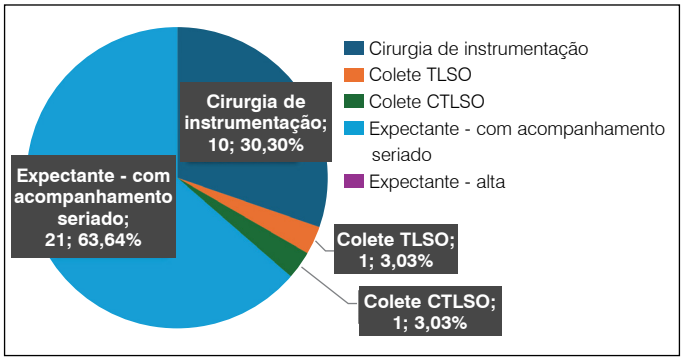

Figura 7. Proporção das respostas dos participantes para o Caso 4.

A média do nível de concordância na escala Likert foi de 2,30 (desvio padrão = 1,076643), com mediana de 2,0 e moda de 2. A Amplitude Interquartil (IQR) foi de 1,0 (Q1 = 2,0; Q3 = 3,0), com variância: 1,159, refletindo maior variabilidade e tendência à discordância com a IA. Não houve correlação significativa entre experiência e concordância (ρ = -0,07; p = 0,63). (Figura 8).

Caso 5

Considerando "colete TLSO" e "colete CTLSO" como tratamento com colete, 6 cirurgiões (18,8%) concordaram com a sugestão da IA. Já a maioria discordou da IA sendo destes: 15 participantes optando por tratamento expectante – com acompanhamento seriado, representando 46,9%, já 11 pariticipantes (34,4%) optaram por tratamento expectante -alta. O coeficiente AC1 de Gwet foi de 0,18, indicando leve concordância. (Figura 9).

A proporção de concordância observada (P_o) entre os cirurgiões no Caso 5 foi de aproximadamente 0,721, indicando concordância substancial. Isso significa que 72,1% das combinações possíveis de respostas entre os cirurgiões foram concordantes. Apesar de haver uma maioria que optou por "expectante", a presença de respostas variadas resultou em uma concordância ligeiramente menor em comparação com o Caso 3.

A média do nível de concordância na escala Likert foi de 1,81 (desvio padrão = 1,221061), com mediana de 1,0 e moda de 1. A Amplitude Interquartil (IQR) foi de 1,0 (Q1 = 1,0; Q3 = 2,0), com variância: 1,49, indicando baixa variabilidade, mas tendência à discordância com a IA foi ainda mais evidente. (Figura 10).

Análise Compilada dos Casos

A Tabela 2 apresenta o padrão de resposta dos especialistas nos casos apresentados e o grau de concordância dos especialistas em relação à resposta do ChatGPT-4.

A Tabela 3 e a Figura 11 apresentam os dados da percepção dos especialistas sobre concordarem com a resposta dada pelo ChatGPT-4 via Escala Likert. Observamos uma diferença estatisticamente significativa entre os casos pelo teste de One-way ANOVA para amostras heterogêneas (F_{welch}=50,7, p<0,001).

O pós-teste de Game-Howell mostra que a percepção de concordância para o Caso 1 é significativamente maior em relação a todos os outros casos (p<0,001) e o Caso 2 apresentou maior percepção de concordância em relação aos Casos 3 (p=0,003) e 5 (p=0,005). O Índice de Validade de Conteúdo (IVC) e o Coeficiente de Validade de Conteúdo (CVC) mostram que apenas no Caso 1 apresentou um valor aceitável entre o nível de concordância dos especialistas com o ChatGPT-4. Valores aceitáveis de concordância são aqueles maiores que 0,8 e preferencialmente devem ser maiores do que 0,9. Calculando o coeficiente Kappa de Fleiss considerando todos os cinco casos, obtivemos um valor de 0,552, indicando uma concordância moderada entre os cirurgiões ao longo dos casos.

Figura 11. Percepção dos especialistas em relação à concordância ao ChatGPT-4 utilizando a escala Likert. Teste realizado: One-way ANOVA (F_{Welch}=50,7, p<0,001). Teste post-hoc de Games-Howell mostra diferenças estatísticas entre o Caso 1 e todos os outros casos (p<0,001) e o Caso 2 em relação aos Casos 3 (p=0,003) e 5 (p=0,005).

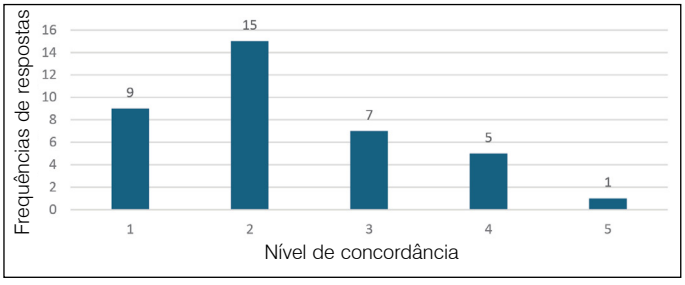

Figura 8. Proporção de respostas da escala Likert para o Caso 4.

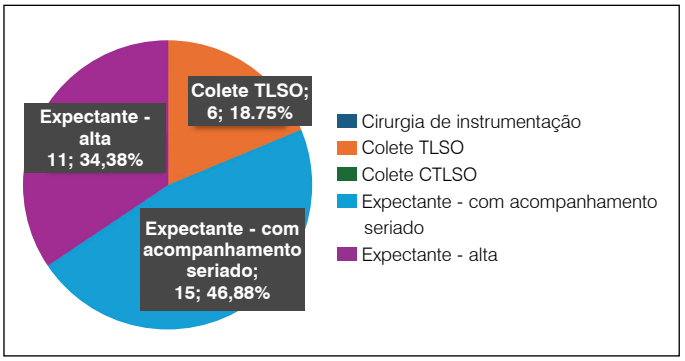

Figura 9. Proporção das respostas dos participantes para o Caso 5.

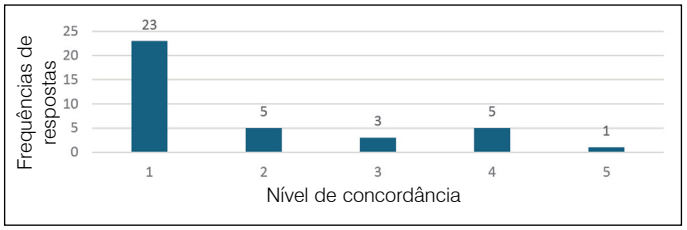

Figura 10. Proporção de respostas da escala Likert para o Caso 5.

Tabela 2. Padrão de resposta dos especialistas nos casos apresentados e porcentagem de concordância entre as soluções de conduta dos especialistas e do ChatGPT-4 (N=37).

	N (%)				
	Caso 1	**Caso 2**	**Caso 3**	**Caso 4**	**Caso 5**
Colete TLSO	1 (2,7%)	22 (61,1%)	3 (8,3%)	1 (3,0%)	6 (18,8%)
Colete CTLSO	0 (0%)	11 (30,6%)	1 (2,8%)	1 (3,0%)	0 (0%)
Cirurgia	36 (97,3%)	1 (2,8%)	32 (88,9%)	10 (30,3%)	0 (0%)
Exp – acomp.	0 (0%)	2 (5,6%)	0 (0%)	21 (63,6%)	15 (46,9%)
Exp - alta	0 (0%)	0 (0%)	0 (0%)	0 (0%)	11 (34,4%)
Total	37 (100%)	36 (100%)	36 (100%)	33 (100%)	32 (100%)
% Concordância c/ ChatGPT-4	97,3% (Surgery)	91,7% (Orthosis)	11,1% (Orthosis)	30,3% (Surgery)	18,8% (Orthosis)

Tabela 3. Percepção dos especialistas em relação à concordância pela escala Likert (N=37).

	Escala Likert						
	n	**IVC**	**CVC**	**Média ± DP**	**Moda**	**Mediana**	**IC (95%)**
Caso 1	37	0,81	0,85	4,3±0,8	5	4	4,01-4,53
Caso 2	36	0,50	0,59	2,9±1,4	4	3,5	2,46-3,43
Caso 3	37	0,05	0,38	1,9±0,9	2	2	1,60-2,18
Caso 4	37	0,16	0,46	2,3±1,1	2	2	1,94-2,66
Caso 5	37	0,16	0,36	1,8±1,2	1	1	1,40-2,22
*P**				<0,001			

* teste de One-way ANOVA para amostras heterogêneas.

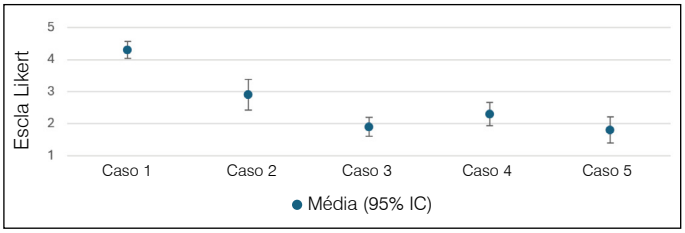

Figura 11. Percepção dos especialistas em relação à concordância ao ChatGPT-4 utilizando a escala Likert. Teste realizado: One-way ANOVA (FWelch=50,7, p<0,001). Teste post-hoc de Games-Howell mostra diferenças estatísticas entre o Caso 1 e todos os outros casos (p<0,001) e o Caso 2 em relação aos Casos 3 (p=0,003) e 5 (p=0,005).

Influência da Experiência Profissional

Não foram encontradas correlações significativas entre a experiência profissional e o nível de concordância com o ChatGPT em nenhum dos casos. Os coeficientes de correlação de Spearman variaram de -0,12 a -0,03, com valores de p acima de 0,05. Isso sugere que a experiência dos cirurgiões não teve um impacto significativo na concordância com as recomendações da IA.

Foi verificada a percepção da concordância pela escala Likert de acordo com os anos de experiência do especialista. Observamos uma diferença estatisticamente significativa apenas na percepção dos especialistas em relação ao Caso 1 (p=0,004) No Caso 1, especialistas com menos de 2 anos de experiência apresentaram valores de média maiores na escala Likert comparado com aqueles com experiência entre 5 e 10 anos (p=0,041) e com mais de 10 anos de experiência (p=0,021) (Tabela 4 e Figura 12).

Tabela 4. Percepção dos especialistas em relação à concordância pela escala Likert por anos de experiência (N=37).

		Escala Likert				
	n	**Caso 1**	**Caso 2**	**Caso 3**	**Caso 4**	**Caso 5**
Menos de 2 anos	9	4,89±0,33	2,56±1,42	1,78±0,67	2,78±0,97	1,78±1,56
Entre 2 a 5 anos	9	4,11±0,93	3,89±0,93	2,33±1,00	2,22±1,09	2,33±1,41
Entre 5 a 10 anos	10	4,20±0,63	2,67±1,58	1,50±0,71	2,30±1,16	1,70±1,16
Mais de 10 anos	9	3,89±0,78	2,67±1,50	2,00±1,00	1,89±1,05	1,44±0,53
*p**		0,004	0,073	0,257	0,370	0,398

* teste de One-way ANOVA para amostras heterogêneas.

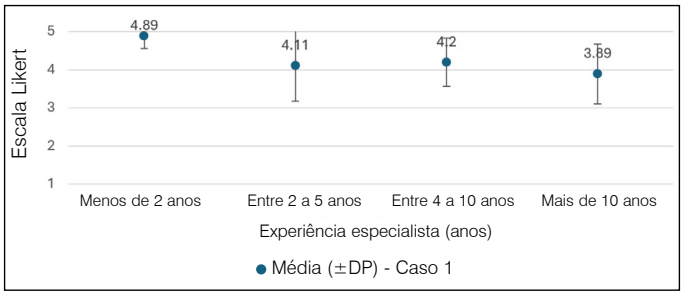

Figura 12. Percepção dos especialistas em relação à concordância ao ChatGPT-4 no Caso 1 utilizando a escala Likert de 5 pontos por anos de experiência do especialista. Teste realizado: One-way ANOVA (F=6,39, p=0,004). Teste post-hoc de Games-Howell mostra diferenças estatísticas entre os especialistas com menos de 2 anos de experiência e aqueles entre 5 e 10 anos (p=0,041) e com mais de 10 anos de experiência (p=0,021).

DISCUSSÃO

Embora tenha acesso a dados médicos limitados, o ChatGPT, incrivelmente, obteve um desempenho equivalente ao de um estudante de medicina do terceiro ano em exames de licenciamento médico largamente aplicados nos Estados Unidos, o que tem estimulado discussões urgentes na área médica.[11]

Segundo Stokel-Walker e van Noorden o ChatGPT poderia responder a algumas perguntas médicas de forma aberta quase tão bem quanto um médico humano médio, embora ainda apresentasse deficiências e imprecisões.[12]

No entanto, apesar das vastas possibilidades, também temos que confrontar as limitações e desafios éticos que surgem com essa tecnologia. Com ênfase especial no potencial para ameaças à segurança de dados e privacidade, o uso adequado e supervisionado do ChatGPT/GPT-4 é de importância primordial, chegando à conclusão de que ChatGPT/GPT-4, quando utilizado com cuidado e responsabilidade, pode ser um farol útil para cirurgiões de coluna 8. Não se deve ignorar as questões de imprecisão dos dados gerados, má conduta e plágio, que demandam atenção.[13]

Na telemedicina, as informações geralmente se limitam à linguagem,[13] tornando-a potencialmente adequada para intervenções usando LLMs. No entanto, o ChatGPT não consegue fazer perguntas para buscar esclarecimentos adicionais sobre perguntas ou cenários.

Baseando-se nas observações de Howard et. al., com base em perguntas feitas ao ChatGPT sobre situações clínicas envolvendo condutas e a tomada de decisão da escolha antimicrobiana, sendo o chatbot capaz de reconhecer a importância de fatores clinicamente relevantes quando informações explícitas foram fornecidas, mas perdeu questões relevantes em cenários de crescente complexidade. A ortografia e gramática nas respostas foram coerentes, e o significado foi claro. Este estudo conclui que os maiores obstáculos à implementação desta IA na prática clínica são as deficiências em consciência situacional e inferência. Essas falhas podem colocar em risco a segurança do paciente. O ChatGPT parece ter acesso a dados de treinamento suficientes, apesar de não ter acesso a bancos de dados médicos específicos. Mesmo sem treinamento específico em aconselhamento clínico, fornece respostas convincentes para a maioria das solicitações.[14]

Os resultados deste estudo fornecem insights importantes sobre a interação entre inteligência artificial e prática clínica no contexto da EIA. O Caso 1, de menor complexidade, apresentou a maior concordância entre os especialistas e o ChatGPT-4 (97%), ao passo que o Casos 3 e 5 apresentaram os menores índices de concordância, sendo 8% e 19% respectivamente.

O ChatGPT demonstrou capacidade de interpretar dados clínicos e fornecer recomendações alinhadas com as práticas estabelecidas em casos simples, como evidenciado pela alta concordância no Caso 1, tanto entre os participantes da pesquisa, quanto em relação à IA. A média elevada na escala Likert (4,27) e os altos coeficientes de concordância (Coeficiente AC1 de Gwet entre os cirurgiões e o ChatGPT foi de 0,95) sugerem um consenso clínico bem estabelecido e confiança na recomendação da IA, para este caso. A proporção de concordância observada (P_o) de 0,946 entre os cirurgiões no Caso 1 indica um consenso clínico sólido. Estudos anteriores corroboram que a IA pode ser eficaz em suportar decisões clínicas padronizadas.[15–17]

Nos casos de maior complexidade (Casos 3 a 5), a concordância diminuiu significativamente. Isso pode ser atribuído à variabilidade nas opções de tratamento em situações clínicas complexas e à necessidade de julgamento clínico refinado. A IA, na sua forma atual, pode não ser capaz de capturar fatores subjetivos e individualizados que influenciam a tomada de decisão, como preferências do paciente, contexto socioeconômico e experiência clínica acumulada.[1,7,18] Por exemplo, no Caso 3, embora a proporção de concordância entre os cirurgiões tenha sido substancial ($P_o \approx 0{,}708$), indicando um consenso entre os profissionais, a alta discordância dos cirurgiões com o ChatGPT ressalta as limitações da IA em contextos complexos.

A concordância geral entre os cirurgiões, indica uma concordância moderada ao longo dos cinco casos. Isso sugere que, embora haja consenso em alguns casos, há variabilidade nas decisões clínicas, possivelmente influenciada pela complexidade dos casos e pelas preferências individuais. Como o exemplo do caso 5, onde houve uma variabilidade de respostas em condutas, com apenas 18,8% dos participantes concordando com a IA.

A ausência de correlação significativa entre a experiência dos cirurgiões e o nível de concordância com o ChatGPT indica que a discordância com a IA não está relacionada ao nível de experiência profissional. Tanto cirurgiões menos experientes quanto os mais experientes demonstraram avaliar criticamente as recomendações da IA, especialmente em casos complexos que exigem julgamento clínico refinado.

Identificaram-se situações em que cirurgiões escolheram a mesma opção de tratamento que o ChatGPT, mas atribuíram baixos níveis de concordância na escala Likert. Isso sugere que, embora a opção de tratamento fosse a mesma, os raciocínios ou detalhes específicos apresentados pela IA não foram totalmente aceitos.

A precisão do ChatGPT na classificação de Lenke evidencia sua habilidade em interpretar informações clínicas objetivas e aplicar critérios complexos de maneira consistente. Isso é particularmente relevante, considerando que a classificação de Lenke é fundamental no planejamento cirúrgico da EIA.[19,20] A capacidade da IA em processar e aplicar classificações padronizadas pode auxiliar na padronização de condutas e reduzir a variabilidade nas decisões clínicas,[21] especialmente com o avanço das tecnologias de Machine learning, considerada uma ferramenta valiosa para planejamento pré-cirúrgico, orientação intraoperatória, monitoramento neurofisiológico e previsão de resultados neurocirúrgicos.[18,21]

As discrepâncias identificadas entre a opção de tratamento e o nível de concordância sugerem que concordar na escolha do tratamento não implica necessariamente concordar nos detalhes ou nos raciocínios subjacentes. Por exemplo, a IA pode recomendar níveis de artrodese diferentes daqueles que os cirurgiões consideram ideais, o que afeta a avaliação geral da recomendação. Além disso, diferenças na interpretação de dados clínicos ou na aplicação de evidências podem levar a discordâncias, mesmo quando a opção de tratamento é a mesma. Ademais, a IA pode ter apresentado conceitos ou detalhes técnicos inadequados que não correspondem à prática clínica atual. Isso ressalta a importância de a IA fornecer não apenas a recomendação, mas também uma explicação clara e fundamentada.[22]

Alguns cirurgiões preferem o TLSO mesmo para curvas acima de T8 devido à melhor adesão ao tratamento, enquanto outros optam pelo CTLSO em conformidade com as recomendações tradicionais. Estudos indicam que o TLSO é eficaz no controle da progressão da escoliose idiopática do adolescente (EIA) e pode ser mais confortável, menos restritivo e com a aparência menos conspícua, o que pode melhorar a adesão ao uso do colete.[23,24]

Considerando que as opções "colete TLSO" e "colete CTLSO" são semelhantes, especialmente no contexto clínico, agrupamos essas respostas como "tratamento com colete". Essa decisão reflete a prática clínica, onde a escolha do tipo de colete pode variar conforme a experiência do cirurgião e as preferências do paciente. A variabilidade nas escolhas de tratamento com colete pode destacar a necessidade de personalização das recomendações da IA 20.

Os achados deste estudo sugerem que o ChatGPT pode ser uma ferramenta auxiliar valiosa no apoio à decisão clínica em casos de EIA com consenso estabelecido. No entanto, em casos complexos, a variabilidade nas opiniões dos cirurgiões destaca a importância do julgamento clínico individual. A IA deve ser vista como uma ferramenta complementar que pode enriquecer o processo de tomada de decisão, mas não substitui a expertise e o raciocínio do profissional de saúde.

Futuras pesquisas devem explorar a integração da IA no fluxo clínico, desenvolvendo estratégias para que ferramentas como o ChatGPT complementem, e não substituam, o raciocínio médico, tornando-as capazes de integrar informações clínicas mais detalhadas, incluindo a satisfação do paciente e a eficácia do tratamento. Além disso, é fundamental trabalhar no aprimoramento dos modelos de IA para que possam considerar fatores individuais do paciente e oferecer recomendações mais personalizadas. A educação e treinamento dos profissionais de saúde no uso dessas tecnologias também são essenciais para maximizar seus benefícios e minimizar potenciais riscos.[25]

Este estudo apresenta algumas limitações. A utilização de casos fictícios pode não refletir plenamente a complexidade e variabilidade dos casos reais. O tamanho da amostra, embora adequado para análises exploratórias, pode limitar a generalização dos resultados.

Nesta jornada de investigação, desbravamos novos caminhos para a prática da medicina e da cirurgia, oferecendo insights sobre como a inteligência artificial, particularmente o ChatGPT/GPT-4, pode se tornar uma ferramenta inestimável. Este trabalho é um marco importante para entender o impacto desta tecnologia avançada na cirurgia da coluna vertebral e moldar o futuro da prática cirúrgica.

CONCLUSÃO

Este estudo evidenciou que o ChatGPT tem capacidade para interpretar dados clínicos e fornece recomendações alinhadas com práticas estabelecidas em casos de EIA. A alta concordância em casos simples indica potencial para uso como suporte à decisão clínica.

No entanto, a variabilidade observada em casos complexos e a influência da experiência profissional ressaltam a importância do julgamento clínico e da individualização do tratamento. A IA deve ser vista como uma ferramenta complementar, que pode enriquecer o processo de tomada de decisão, mas não substitui a expertise e o raciocínio do profissional de saúde. Especialmente no contexto atual, onde ainda ocorrem muitos erros de interpretação de conceitos e de priorização de dados clínicos.

A integração eficaz da IA na prática clínica requer compreensão de suas limitações e potencialidades, bem como esforços contínuos para aprimorar suas aplicações no contexto médico.

All authors declare no potential conflict of interest related to this article.

CONTRIBUIÇÕES DOS AUTORES: Cada autor contribuiu individual e significantemente para o desenvolvimento deste artigo. LSRO: Conceitualização, Metodologia, aquisição de dados e Redação do trabalho; RS: Contribuição substancial na concepção, análise formal e aquisição de dados; AC: Contribuição substancial na concepção, interpretação dos dados, Aquisição de financiamento; BM: Curadoria de dados, análise e aquisição dos dados; WO: Metodologia e aquisição dos dados; GF: interpretação dos dados e revisão crítica do seu conteúdo intelectual; GAVP: Contribuição substancial na concepção, interpretação dos dados; Gabrielle do Amaral Virginio Pereira: Contribuição substancial na concepção, interpretação dos dados; LRR: Validação e escrita – revisão e edição, Aprovação final da versão do manuscrito a ser publicado.

REFERÊNCIAS

1. Sohail SS. A Promising Start and Not a Panacea: ChatGPT's Early Impact and Potential in Medical Science and Biomedical Engineering Research. Ann Biomed Eng. 2024;52(5):1131-1135. doi: 10.1007/s10439-023-03335-6.
2. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. IBM J Res Dev. 1959;3(3):210–29. doi: 10.1147/rd.33.0210.
3. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, et al. Deep Learning in Medical Imaging: General Overview. Korean J Radiol. 2017;18(4):570-584. doi: 10.3348/kjr.2017.18.4.570.
4. King BF Jr. Artificial Intelligence and Radiology: What Will the Future Hold? J Am Coll Radiol. 2018;15(3 Pt B):501-503. doi: 10.1016/j.jacr.2017.11.017.
5. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. Insights Imaging. 2018;9(5):745-753. doi: 10.1007/s13244-018-0645-y.
6. Zhang N, Sun Z, Xie Y, Wu H, Li C. The latest version ChatGPT powered by GPT-4o: what will it bring to the medical field? Int J Surg. 2024;110(9):6018-6019. doi: 10.1097/JS9.0000000000001754.
7. Bečulić H, Begagić E, Skomorac R, Mašović A, Selimović E, Pojskić M. ChatGPT's contributions to the evolution of neurosurgical practice and education: a systematic review of benefits, concerns and limitations. Med Glas (Zenica). 2024;21(1). doi: 10.17392/1661-23.
8. He Y, Tang H, Wang D, Gu S, Ni G, Wu H. Will ChatGPT/GPT-4 be a Lighthouse to Guide Spinal Surgeons? Ann Biomed Eng. 2023;51(7):1362-1365. doi: 10.1007/s10439-023-03206-0.
9. The jamovi project. jamovi. [Internet]. 2021. Retrieved from: https://www.jamovi.org/.
10. R Core Team. A Language and environment for statistical computing. [Internet]. 2021. Retrieved from: https://search.gesis.org/publication/zis-RCoreTeam.2021R.
11. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. JMIR Med Educ. 2023;9:e45312. doi: 10.2196/45312.
12. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. Nature. 2023;614(7947):214-216. doi: 10.1038/d41586-023-00340-6.
13. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 Pandemic, Lessons for the Future. Telemed J E Health. 2020;26(5):571-573. doi: 10.1089/tmj.2020.29040.rb.
14. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? Lancet Infect Dis. 2023;23(4):405-406. doi: 10.1016/S1473-3099(23)00113-5.
15. Ouanes K, Farhah N. Effectiveness of Artificial Intelligence (AI) in Clinical Decision Support Systems and Care Delivery. J Med Syst. 2024;48(1):74. doi: 10.1007/s10916-024-02098-4.
16. Lüscher TF, Wenzl FA, D'Ascenzo F, Friedman PA, Antoniades C. Artificial intelligence in cardiovascular medicine: clinical applications. Eur Heart J. 2024;45(40):4291-4304. doi: 10.1093/eurheartj/ehae465.
17. Han C, Pan Y, Liu C, Yang X, Li J, Wang K, et al. Assessing the decision quality of artificial intelligence and oncologists of different experience in different regions in breast cancer treatment. Front Oncol. 2023;13:1152013. doi: 10.3389/fonc.2023.1152013.
18. Senders JT, Zaki MM, Karhade A V., Chang B, Gormley WB, Broekman ML, et al. An introduction and overview of machine learning in neurosurgical care. Acta Neurochir (Wien). 2018;160(1):29-38. doi: 10.1007/s00701-017-3385-8.
19. Lenke LG, Edwards CC 2nd, Bridwell KH. The Lenke classification of adolescent idiopathic scoliosis: how it organizes curve patterns as a template to perform selective fusions of the spine. Spine (Phila Pa 1976). 2003;28(20):S199-207. doi: 10.1097/01.BRS.0000092216.16155.33.
20. Hoashi JS, Cahill PJ, Bennett JT, Samdani AF. Adolescent scoliosis classification and treatment. Neurosurg Clin N Am. 2013;24(2):173-83. doi: 10.1016/j.nec.2012.12.006.
21. Zhou S, Zhou F, Sun Y, Chen X, Diao Y, Zhao Y, et al. The application of artificial intelligence in spine surgery. Front Surg. 2022;9:885599. doi: 10.3389/fsurg.2022.885599.
22. Yang Q, Steinfeld A, Zimmerman J. Unremarkable AI. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM. 2019;(238):1–11. doi.org/10.1145/3290605.3300468
23. Howard A, Wright JG, Hedden D. A comparative study of TLSO, Charleston, and Milwaukee braces for idiopathic scoliosis. Spine (Phila Pa 1976). 1998;23(22):2404-11. doi: 10.1097/00007632-199811150-00009.
24. Janicki JA, Poe-Kochert C, Armstrong DG, Thompson GH. A comparison of the thoracolumbosacral orthoses and providence orthosis in the treatment of adolescent idiopathic scoliosis: results using the new SRS inclusion and assessment criteria for bracing studies. J Pediatr Orthop. 2007;27(4):369-74. doi: 10.1097/01.bpb.0000271331.71857.9a.
25. Tekkeşin Aİ. Artificial Intelligence in Healthcare: Past, Present and Future. Anatol J Cardiol. 2019;22(Suppl 2):8-9. doi: 10.14744/AnatolJCardiol.2019.28661.