



Avaliação da qualidade das respostas do modelo de inteligência artificial ChatGPT 4 no plano terapêutico de pacientes obesos

Evaluation of the quality of responses from the ChatGPT 4 artificial intelligence model in the therapeutic plan of obese patients

Evaluación de la calidad de las respuestas del modelo de inteligencia artificial ChatGPT 4 en el plan terapéutico de pacientes obesos

Louise Menezes da Cunha¹, Victor Menezes da Cunha¹, Rafael Melillo Laurino Neto¹, Alecsander Rodriguez Ojea¹, Larissa Sena Cotrim¹, Heron Kairo Saboia Sant'Anna Lima¹, Mariana Costa¹, Roberta Tsuboi Rodrigues², Ana Victoria Martins Lima³, Matheus de Sousa Pantoja³.

RESUMO

Objetivo: Avaliar a qualidade das recomendações fornecidas pelo modelo de inteligência artificial ChatGPT 4 acerca do manejo de pacientes obesos em casos clínicos fictícios. **Métodos:** Trata-se de um estudo não experimental, de caráter observacional e transversal, com abordagem quantitativa, em que cirurgiões bariátricos analisaram a qualidade das respostas fornecidas pelo modelo de IA (ChatGPT 4) acerca do plano terapêutico, indicação cirúrgica e técnica cirúrgica caso indicado. **Resultados:** O ChatGPT 4 apresentou uma concordância média de 54,2% com as opiniões dos especialistas, sendo a identificação da elegibilidade para cirurgia mais alta (80% dos casos). Os cirurgiões pontuaram o plano terapêutico com escore médio de 76,4. No entanto, houve significativa discordância na escolha da técnica operatória, principalmente em situações que exigiam julgamento clínico complexo. **Conclusão:** Este trabalho conclui que, embora o ChatGPT 4 mostre-se promissor, sua utilização precisa ser aprimorada mediante uma abordagem colaborativa que integre conhecimento humano e tecnologia. **Conclusão:** As respostas do ChatGPT 4 foram classificadas como “totalmente adequadas” em pouco mais da metade das avaliações, sugerindo que, embora a IA possua um potencial promissor para auxiliar na tomada de decisão clínica, ainda necessita de refinamentos para atingir um nível de confiabilidade compatível com a prática médica.

Palavras-chave: Obesidade, Inteligência Artificial, Cirurgia bariátrica.

ABSTRACT

Objective: To assess the quality of the recommendations provided by the artificial intelligence model ChatGPT 4 regarding the management of obese patients in fictional clinical cases. **Methods:** This is a non-experimental, observational, and cross-sectional study with a quantitative approach, in which bariatric surgeons analyzed the quality of the responses provided by the AI model (ChatGPT 4) concerning the therapeutic plan, surgical indication, and surgical technique when applicable. **Results:** ChatGPT 4 showed an average agreement rate of 54.2% with expert opinions, with the identification of surgical eligibility being the highest (80% of cases).

¹ Conjunto Hospitalar do Mandaqui (CHM), São Paulo - SP.

² Hospital Emílio Ribas, São Paulo - SP.

³ Santa Casa de Misericórdia, São Paulo - SP.

The surgeons rated the therapeutic plan with an average score of 76.4. However, there was significant disagreement in the choice of surgical technique, particularly in situations requiring complex clinical judgment.

Conclusion: This study concludes that while ChatGPT 4 appears promising, its use needs improvement through a collaborative approach integrating human knowledge and technology. ChatGPT 4's responses were classified as "fully adequate" in just over half of the evaluations, suggesting that although AI has promising potential to assist in clinical decision-making, it still requires refinement to achieve a level of reliability compatible with medical practice.

Keywords: Obesity, Artificial Intelligence, Bariatric surgery.

RESUMEN

Objetivo: Evaluar la calidad de las recomendaciones proporcionadas por el modelo de inteligencia artificial ChatGPT 4 sobre el manejo de pacientes obesos en casos clínicos ficticios. **Métodos:** Se trata de un estudio no experimental, de carácter observacional y transversal, con un enfoque cuantitativo, en el que cirujanos bariátricos analizaron la calidad de las respuestas proporcionadas por el modelo de IA (ChatGPT 4) en relación con el plan terapéutico, la indicación quirúrgica y la técnica quirúrgica cuando fue indicada.

Resultados: ChatGPT 4 mostró un nivel medio de concordancia del 54,2 % con las opiniones de los especialistas, siendo la identificación de la elegibilidad para la cirugía la más alta (80 % de los casos). Los cirujanos calificaron el plan terapéutico con un puntaje medio de 76,4. Sin embargo, hubo una discrepancia significativa en la elección de la técnica quirúrgica, especialmente en situaciones que requerían un juicio clínico complejo. **Conclusión:** Este estudio concluye que, aunque ChatGPT 4 muestra un potencial prometedor, su uso debe mejorarse mediante un enfoque colaborativo que integre el conocimiento humano y la tecnología. Las respuestas de ChatGPT 4 fueron calificadas como "totalmente adecuadas" en poco más de la mitad de las evaluaciones, lo que sugiere que, si bien la IA tiene un potencial prometedor para ayudar en la toma de decisiones clínicas, aún necesita refinamientos para alcanzar un nivel de fiabilidad compatible con la práctica médica.

Palabras clave: Obesidad, Inteligencia Artificial, Cirugía bariátrica.

INTRODUÇÃO

A cirurgia bariátrica consolidou-se como uma das intervenções mais eficazes no tratamento da obesidade. Além da perda de peso substancial, essa abordagem contribui para melhorias na qualidade de vida dos pacientes, reduzindo a mortalidade, proporcionando benefícios metabólicos duradouros e redução significativa de comorbidades como diabetes tipo 2, hipertensão e apneia do sono. O sucesso desse procedimento cirúrgico depende de uma cuidadosa avaliação pré-operatória, destinada a identificar fatores de risco, otimizar condições clínicas e selecionar candidatos para a cirurgia.

Essa avaliação, tradicionalmente realizada por uma equipe multidisciplinar – composta por cirurgiões, endocrinologistas, cardiologistas, nutricionistas e psicólogos, entre outros – garante uma análise completa de cada aspecto da saúde do paciente (JAZI AHD, et al., 2023; COURCOULAS AP, et al., 2023; ASKARI A, et al., 2023; PERDOMO CM, et al., 2023; RAJEEV ND, et al., 2023; PREMKUMAR A, et al., 2022).

Nos últimos anos, observa-se um fenômeno relevante: a crescente busca de pacientes por informações na internet antes de decidir por procedimentos invasivos como a cirurgia bariátrica. Estudos indicam que 77,8% dos pacientes consideram a internet uma fonte primária para obter informações de saúde, o que pode tanto empoderá-los quanto expô-los a informações de qualidade variável (JAZI AHD, et al., 2023). Essa tendência reflete o papel crescente da internet na prática médica e sublinha a necessidade de garantir que as informações acessadas sejam confiáveis e baseadas em evidências (LEE Y, et al., 2024; SRINIVASAN N, et al., 2024; TARIQ R e KHANNA S, 2023).

Diante dessa realidade, a inteligência artificial (IA) surge como uma alternativa promissora para fornecer informações de saúde personalizadas e acessíveis. Modelos de IA baseados em diálogos, como o ChatGPT 4, apresentam-se como ferramentas potenciais para atender a essa demanda, oferecendo informações

relevantes para pacientes e profissionais de saúde. Os modelos de IA têm a capacidade de processar grandes volumes de dados médicos em tempo real, o que os posiciona como recursos auxiliares para esclarecer opções de tratamento e responder a dúvidas em saúde (LEE Y, et al., 2024; LOPEZ-GONZALEZ R, et al. 2024; JAZI AHD, et al., 2023; TARIQ R e KHANNA S, 2023).

Entre essas tecnologias, destaca-se o ChatGPT 4, desenvolvido pela OpenAI e lançado em março de 2023. Essa versão traz melhorias expressivas em relação às suas predecessoras, especialmente no que se refere à compreensão e análise de cenários clínicos complexos (JAZI AHD, et al., 2023). A capacidade do ChatGPT 4 de simular interações em linguagem natural o torna uma ferramenta promissora tanto para a educação em saúde quanto para o suporte na tomada de decisões clínicas. Estudos recentes sugerem que o ChatGPT 4 possui a capacidade de responder com precisão a exames médicos e avaliações clínicas de alta complexidade, o que evidencia seu potencial para aplicação em diferentes contextos cirúrgicos (OH N, et al., 2023; ALI A, et al., 2023; LONG C, et al., 2023; MUSTAFA A, et al., 2023).

Esses estudos indicam que este modelo de IA pode alcançar um nível de competência comparável ao de profissionais de saúde experientes, inclusive em questões que exigem conhecimento técnico avançado (TEMSAH O, et al., 2023). Contudo, é importante reconhecer que, apesar das promessas, a aplicação da IA na prática clínica enfrenta grandes desafios (TEMSAH O, et al., 2023). O ChatGPT 4 ainda possui limitações na interpretação de dados clínicos complexos e podem não captar nuances que especialistas humanos detectariam, como fatores psicossociais que influenciam as decisões dos pacientes e sua adesão ao tratamento (PERDOMO CM, et al., 2023; RAJEEV ND, et al., 2023; PREMKUMAR A, et al., 2022).

Além disso, a confiabilidade das respostas geradas por esses modelos depende da qualidade dos dados utilizados em seu treinamento. Questões éticas também emergem, especialmente em relação à privacidade dos dados e à responsabilidade em caso de recomendações inadequadas (JAZI AHD, et al., 2023; COURCOULAS AP, et al., 2023; ASKARI A, et al., 2023).

Nesse contexto, este estudo visa avaliar a adequação e consistência das respostas fornecidas pelo ChatGPT 4 em casos clínicos de pacientes obesos, com o objetivo de observar sua viabilidade como ferramenta na avaliação pré-operatória de candidatos à cirurgia bariátrica.

MÉTODOS

Caracterização do Estudo:

Trata-se de um estudo não experimental, de caráter observacional e transversal, com abordagem quantitativa, em que cirurgiões bariátricos analisaram a qualidade das respostas fornecidas pelo modelo de IA (ChatGPT-4) acerca do plano terapêutico, com ênfase na abordagem cirúrgica, para casos clínicos fictícios de pacientes com obesidade.

O estudo foi realizado nas seguintes etapas: (1) Elaboração de cenários clínicos fictícios, (2) Submissão dos cenários ao modelo de IA: Apresentação dos cenários clínicos ao ChatGPT 4, seguida da realização de perguntas previamente estabelecidas (3) Seleção de cirurgiões bariátricos (4) Avaliação das respostas pela equipe de especialistas (5) Análise dos dados

Elaboração de cenários clínicos fictícios

Para este estudo, foram elaborados cinco cenários clínicos fictícios, que contemplaram diferentes aspectos do tratamento da obesidade. Nenhum dado pessoal de pacientes reais foi utilizado, e todos os cenários clínicos empregados foram desenvolvidos exclusivamente para esta pesquisa.

A elaboração destes cenários foi cuidadosamente planejada, buscando um equilíbrio na complexidade, de modo a evitar tanto casos excessivamente simples quanto excessivamente complicados. Assim, os cenários clínicos simularam situações realistas com um nível moderado de dificuldade, proporcionando um ambiente propício para a tomada de decisão clínica.

Questionamentos à IA (ChatGPT 4)

Seleção dos especialistas com critérios de inclusão e exclusão

Foram convidados 30 cirurgiões bariátricos para participar do estudo. O tamanho da amostra de especialistas foi definido com base na literatura vigente, alinhando-se à média de participantes adotada em estudos de metodologia semelhante, que variou de 5 a 30 especialistas (JAZI AHD, et al., 2023; CHIESA-ESTOMBA CM, et al., 2024; GUMMADI R, et al., 2024). Foram incluídos no estudo cirurgiões bariátricos qualificados, que possuem título de especialista Sociedade Brasileira de Cirurgia Bariátrica e Metabólica (SBCBM), com pelo menos 10 anos de experiência na área e que são familiarizados com diferentes técnicas de cirurgias bariátricas. Foram excluídos do estudo os cirurgiões que se recusarem a assinar o Termo de Consentimento Livre e Esclarecido (TCLE) e os que não enviarem as respostas do questionário dentro do prazo estipulado de 15 dias.

A seleção dos especialistas foi realizada por meio de uma busca de currículos na Plataforma Lattes, utilizando a ferramenta de "busca por assunto" e aplicando filtros para profissionais graduados com as palavras-chave "cirurgia bariátrica" e "obesidade". Ademais, foi utilizada a técnica de bola de neve, uma técnica não probabilística de seleção amostral, em que os participantes iniciais indicam novos participantes que atendem aos critérios de inclusão e assim sucessivamente. Essa técnica é bastante utilizada quando a população é composta por pessoas com características difíceis de serem encontradas.

Coleta de Dados

O instrumento de coleta de dados foi dividido em duas etapas. Na primeira etapa do questionário, foram coletadas informações para caracterizar a amostra de especialistas. Na segunda etapa, os cirurgiões tiveram acesso aos cinco cenários clínicos fictícios, acompanhados das respostas fornecidas pelo ChatGPT-4. Essas respostas foram avaliadas pelos cirurgiões através de uma escala Likert, considerando a adequação de três aspectos principais que chamaremos de domínios: plano terapêutico, indicação de cirurgia bariátrica e técnica operatória selecionada. A escala variou de 1 a 3, onde 1 representa respostas "totalmente inadequadas", 2 "parcialmente adequadas," e 3 "totalmente adequadas".

Análise de Dados

Os dados epidemiológicos e pontuações atribuídas pelos especialistas a cada um dos três domínios avaliados nos cenários clínicos foram organizados e processados, através da tabulação em uma planilha eletrônica no software Microsoft Excel. Para caracterização demográfica dos especialistas foi realizada análise descritiva utilizando estatísticas de frequências absolutas e relativas. Na avaliação da adequação das recomendações do ChatGPT, cada resposta foi pontuada de acordo com a escala Likert de 1 a 3. Para facilitar a interpretação dos resultados, os dados coletados foram normalizados, permitindo uma análise mais clara e comparável entre os diferentes especialistas e cenários avaliados.

Essa normalização permitiu identificar padrões nas respostas e a visualizar eventuais divergências ou concordâncias na avaliação dos domínios propostos. A normalização dos dados foi realizada utilizando a **min-max normalization**, transformando os valores da escala Likert (1 a 3) em uma escala padronizada entre 0 e 100. Essa técnica mantém a distribuição original dos dados, preservando as relações entre as avaliações feitas pelos especialistas. A fórmula aplicada foi:

$$X' = \frac{(X - X_{\min})}{(X_{\max} - X_{\min})} \times 100$$

Onde X é o valor original da resposta, Xmin e Xmax correspondem aos valores mínimo e máximo da escala Likert (1 e 3, respectivamente). Assim, as respostas "totalmente inadequadas" (1) serão transformadas em 0, as "parcialmente adequadas" (2) em 50 e as "totalmente adequadas" (3) em 100, gerando um escore.

Aspectos éticos

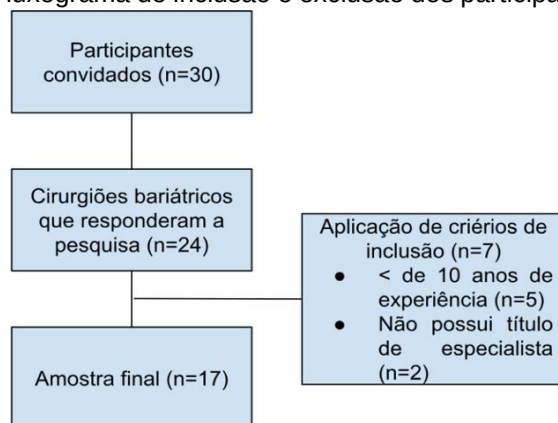
O projeto de pesquisa foi submetido à apreciação e aprovação do Comitê de Ética e Pesquisa (CEP) do Conjunto Hospitalar do Mandaqui, via Plataforma Brasil, de acordo com as normas e preceitos éticos estabelecidos na Resolução 466/2012, que versa sobre as Diretrizes e Normas Regulamentadoras de Pesquisa envolvendo seres humanos. Aprovado sob o parecer de número 7.367.857 e CAAE: 85883724.4.0000.5551. Para participar do estudo, o médico especialista foi orientado sobre os objetivos da

pesquisa, bem como seus riscos e benefícios. Tais informações constam no TCLE, assinado por cada participante antes de iniciar o preenchimento dos questionários.

RESULTADOS

Foram convidados 30 cirurgiões bariátricos para participar da pesquisa, destes 6 cirurgiões não responderam ao convite ou se recusaram a participar, desta forma, 24 cirurgiões bariátricos responderam ao protocolo de pesquisa (**Figura 1**). Destes 7 não preencheram os critérios de inclusão (tempo de experiência) e foram excluídos da análise final, restando uma amostra final de 17 participantes. Dentre os participantes, 94,1% eram do sexo masculino e 5,9% do sexo feminino, com média de idade de 52,3 anos. A média de tempo de experiência na especialidade foi de 19,0 anos (DP $\pm$ 5,22 anos), variando de 13 até 30 anos de experiência.

Figura 1 – Fluxograma de inclusão e exclusão dos participantes do estudo.



Fonte: Cunha VM, et al., 2025.

Quando perguntados sobre quais técnicas de cirurgia bariátrica os cirurgiões se sentiam capazes de realizar, todos responderam ser capazes de realizar pelo menos duas técnicas cirúrgicas, sendo o Sleeve e Bypass Gástrico citado por todos eles. As técnicas menos citadas pelos cirurgiões bariátricos foram as técnicas de bipartição e Sadis, ambas citadas por apenas 2 cirurgiões.

Tabela 1 – Dados epidemiológicos dos cirurgiões bariátricos que participaram da pesquisa.

Dados epidemiológicos	N (média)	% (Desvio padrão)
Sexo		
Masculino	16	94,1%
Feminino	1	5,9%
Idade	52,3	9,41
Tempo de experiência (em anos)	19,0	5,22

Fonte: Cunha VM, et al., 2025.

As respostas fornecidas pelo modelo ChatGPT 4 foram avaliadas em três domínios: (1) à adequação do plano terapêutico, (2) a indicação cirúrgica e (3) a técnica operatória sugerida para cada um dos cinco cenários clínicos fictícios. Para facilitar a discussão e entendimento dos resultados, as respostas da IA foram tabuladas de forma esquemática e resumida na (**Tabela 2**).

Tabela 2 – Resumo dos casos clínicos fictícios apresentados ao ChatGPT, e técnicas cirúrgicas indicadas pela IA.

Cenário	Resumo clínico	Técnica cirúrgica indicada
Caso clínico 1	IMC 34, hipertensão e <i>H. pylori</i> +	Sleeve
Caso clínico 2	IMC 37, tabagista, dislipidêmico e com metaplasia em antro	Sleeve
Caso clínico 3	IMC 28, apneia do sono	Não indicado
Caso clínico 4	IMC 40, hipertensão e refluxo	Sleeve
Caso clínico 5	IMC 37, crohn	Não indicado

Fonte: Cunha VM, et al., 2025.

Os escores médios para cada um dos domínios avaliados estão apresentados na tabela abaixo, com o score médio obtido para cada cenário. Conforme descrito na metodologia, escores mais próximos de zero indicam menor concordância com a indicação da IA, enquanto valores mais altos, próximos de 100, refletem maior concordância.

Tabela 3 – Escore de concordância dos cirurgiões com as respostas do ChatGPT .

Cenário	Plano terapêutico	Indicação de cirurgia bariátrica	Técnica operatória
Caso clínico 1	70,5	67,5	67,5
Caso clínico 2	76,0	79,0	64,5
Caso clínico 3	91,0	97,0	-
Caso clínico 4	62,5	67,5	2,5
Caso clínico 5	82,0	61,5	-

Fonte: Cunha VM, et al., 2025.

A avaliação do plano terapêutico alcançou score médio global (todos os casos) de 76,4. O escore mais baixo no plano terapêutico foi registrado no caso clínico 4, onde apenas 5 especialistas (29,4%) consideraram a resposta totalmente adequada. Enquanto isso, no cenário 3, o ChatGPT apresentou o plano terapêutico com maior taxa de concordância entre os especialistas. Neste cenário 88,2% dos cirurgiões consideraram totalmente adequado o plano terapêutico fornecido pela IA.

A capacidade do ChatGPT 4 em identificar corretamente a necessidade de intervenção cirúrgica obteve um score médio de 74,5. A técnica cirúrgica sugerida pelo modelo foi avaliada com escore médio global de 44,8. O maior nível de discordância ocorreu no Cenário 4, onde nenhum dos cirurgiões considerou totalmente adequada a escolha do ChatGPT-4, e apenas um cirurgião considerou a técnica escolhida como parcialmente adequada.

DISCUSSÃO

O uso de internet para buscar informações em saúde é uma prática cada vez mais comum entre os pacientes, e tópicos como cirurgia bariátrica e emagrecimento são cada vez mais frequentes. No entanto, diversos estudos revelaram que as fontes online relacionadas à cirurgia bariátrica frequentemente carecem de qualidade, precisão e confiabilidade. Esse déficit desafia a educação em saúde e confunde os pacientes, ao fornecer informações incompletas ou enganosas (JAZI AHD, et al., 2023).

Recentemente, o ChatGPT demonstrou sua capacidade de fornecer respostas sobre cirurgia bariátrica, tornando-se uma ferramenta potencialmente valiosa para aprimorar a educação dos pacientes, preencher lacunas de informação e apoiar profissionais de saúde durante consultas. Com base nesse potencial, este estudo avaliou o modelo de IA ChatGPT 4, uma versão avançada do sistema, para determinar sua capacidade de identificar a elegibilidade de pacientes para cirurgia bariátrica e a técnica cirúrgica apropriada. As recomendações da IA foram comparadas às opiniões de um painel de especialistas de diversas regiões do Brasil.

A amostra final deste estudo foi composta por 17 cirurgiões bariátricos que atenderam aos critérios de inclusão estabelecidos. A distribuição por sexo revelou um predomínio masculino (94,1%), evidenciando uma característica frequentemente observada na cirurgia bariátrica e do aparelho digestivo. Dados da última demografia médica evidenciam 87,6% de cirurgiões do sexo masculino, com idade média de 47,2 anos e 23,1 anos de experiência (SCHEFFER M, et al., 2023). A literatura aponta que a participação feminina em especialidades cirúrgicas ainda é reduzida, possivelmente devido a barreiras históricas e desafios inerentes à carreira cirúrgica, como carga horária extensa e demandas físicas e emocionais (PAULO D, et al., 2020).

A média de idade dos participantes foi de 52,3 anos, o que reflete um grupo com vasta experiência profissional. O tempo médio de atuação na especialidade foi de 19 anos, com variação entre 13 e 30 anos, indicando que os profissionais consultados possuem um tempo de prática significativo, o que agrega robustez às respostas coletadas.

Essa experiência substancial é fundamental para a análise das respostas geradas pela IA, uma vez que profissionais experientes tendem a ter uma visão mais consolidada sobre padrões de conduta e tomada de

decisão clínica. Neste estudo, todos os cirurgiões relataram estar aptos a realizar pelo menos duas técnicas cirúrgicas, sendo o Sleeve e o Bypass Gástrico as técnicas mais frequentemente mencionadas.

Esse resultado é coerente com a literatura, que aponta essas técnicas como as mais empregadas no Brasil e no mundo para o tratamento da obesidade (NUNES BC, et al., 2024; OZSOY Z e DEMIR E, 2018). Estudos vêm explorando o potencial da IA, em especial do ChatGPT, no contexto de educação em saúde, especialmente sua capacidade de auxiliar no processo de decisão em cirurgia bariátrica (LOPEZ-GONZALEZ R, et al., 2024; JAZI AHD, et al., 2023). Neste estudo, ao comparar as recomendações do ChatGPT 4 com as opiniões do painel de especialistas, observamos uma taxa de concordância de aproximadamente 54,2%, um índice consideravelmente superior ao relatado em estudos anteriores, cuja média de concordância era em torno de 30% (LOPEZ-GONZALEZ R, et al., 2024; JAZI AHD, et al., 2023).

Em outro estudo recente, a IA respondeu corretamente 90% das perguntas sobre o tema (SAMAAN, et al. 2023), demonstrando a relevância do conhecimento da IA acerca da cirurgia bariátrica. Essa crescente aproximação entre as recomendações da IA e a boa prática médica pode indicar uma evolução progressiva dos modelos de IA refletindo aprimoramentos na qualidade e precisão das respostas fornecidas. Esses resultados sugerem que, embora a IA não possa substituir o julgamento clínico humano, ela pode se tornar uma ferramenta valiosa para complementar o processo de decisão e auxiliar na padronização de condutas baseadas em evidências.

Quando avaliamos a capacidade da IA em identificar candidatos adequados para cirurgia bariátrica, o ChatGPT 4 recomendou a cirurgia em 60% dos casos, enquanto o painel de especialistas indicou procedimento cirúrgico 64,7% das vezes, exibindo um nível de concordância de 88,3% entre o painel de especialistas e o ChatGPT-4. Outras investigações também identificaram elevada taxa de sucesso em identificar pacientes com indicação cirúrgica, com taxas superiores a 80% de assertividade (LOPEZ-GONZALEZ R, et al., 2024; PANTELIS A, 2023).

Apesar do bom desempenho da IA em identificar a necessidade de intervenção cirúrgica, a seleção da técnica operatória apresentou maior variabilidade, refletindo sua limitação em captar fatores clínicos mais sutis, como a presença de comorbidades específicas. Ao avaliar as escolhas da técnica cirúrgica pelo ChatGPT 4, observou-se discrepância significativa comparada à escolha dos especialistas, especialmente no caso 4.

Neste caso, avaliou-se um paciente com IMC operável e doença do refluxo grave, para o qual a IA indicou a técnica de sleeve, uma escolha amplamente desaconselhada na literatura devido ao potencial de agravamento do refluxo (LOPES D, 2024). Como esperado, essa conduta foi rejeitada pela maioria dos especialistas. Situação semelhante ocorreu no estudo realizado recentemente (JAZI AHD, et al., 2023), em que a IA falhou em indicar a técnica cirúrgica adequada em um paciente doença do refluxo gastro-esofágico. Esse achado pode indicar uma limitação da IA em considerar complicações a longo prazo ou refletir deficiências em sua base de dados, o que levou à indicação do sleeve de forma recorrente, mesmo quando existiam opções cirúrgicas mais apropriadas. A insistência nesse padrão reforça a necessidade de uma abordagem crítica ao utilizar a IA antes de sua aplicação na prática médica.

Essa divergência é um dos exemplos que reforçam a importância da supervisão humana ao utilizar IA no suporte à decisão clínica, destacando a necessidade de aprimoramentos no modelo para incorporar critérios mais refinados e alinhados às diretrizes médicas atuais. Esses resultados destacam tanto o potencial quanto às limitações do ChatGPT 4 como uma ferramenta de apoio à decisão em cirurgia bariátrica, enfatizando a necessidade de aprimoramentos adicionais para melhorar sua confiabilidade e aplicabilidade clínica.

Nos dois primeiros cenários (Caso clínico 1 e 2) as respostas da IA alcançaram elevado escore de concordância com os especialistas, em todos os aspectos avaliados (>60). Estes casos abordaram pacientes com obesidade de grau variado, associado a uma ou mais comorbidades (*H. pylori* resistente com falha na perda ponderal com dieta e exercício, e , metaplasia em antro). Em ambas as situações a IA sugeriu tratamento clínico com medicações antiobesidade e, em caso de insucesso, cirurgia metabólica (Sleeve).

O desempenho favorável da IA em cenários menos complexos é consistente com achados de estudos prévios (LOPEZ-GONZALEZ R, et al., 2024; JAZI AHD, et al., 2023), sugerindo que, quando os casos seguem diretrizes bem estabelecidas, a IA tende a oferecer recomendações alinhadas com a prática clínica. Os casos

3 e 5 abordam pacientes com sobrepeso e obesidade grau I, sem quaisquer indicações cirúrgicas. A concordância com a IA nesses casos foi elevada, com escores acima de 90. Esse resultado pode refletir a baixa complexidade do caso. O caso 5 aborda um paciente com obesidade grau 1, sem comorbidades associadas à obesidade e doença inflamatória intestinal. Não foi indicada abordagem cirúrgica pela IA. Houve mínima discordância de alguns especialistas.

A avaliação dos achados do estudo reforça tanto o potencial quanto às limitações do ChatGPT 4 como ferramenta auxiliar na decisão cirúrgica, demonstrando que, embora ela possa contribuir para padronizar condutas e agilizar processos, sua implementação na prática clínica ainda requer ajustes e validação constante. A aceitação dessa tecnologia pelos profissionais de saúde dependerá não apenas de sua precisão, mas também da segurança proporcionada ao paciente. Assim, a evolução dos modelos de IA para aplicações clínicas deve caminhar lado a lado com a supervisão humana e a análise crítica, garantindo que tais ferramentas complementam, e não substituem, o julgamento médico qualificado.

Este estudo apresenta limitações inerentes à metodologia adotada. O uso de cenários clínicos fictícios, embora cuidadosamente elaborados para refletir a prática real, pode não captar todas as nuances presentes nos atendimentos reais, especialmente no que se refere a fatores psicossociais dinâmicos. Outra limitação importante da IA, já descrita na literatura, é a inconsistência observada nas respostas do ChatGPT 4. Estudos relatam que, quando questionado sobre o mesmo caso em momentos distintos, o modelo pode apresentar uma variabilidade de até 40% em suas respostas (JAZI AHD, et al., 2023).

Essa inconsistência não foi confirmada nesta pesquisa. Outro fator relevante é a subjetividade nas avaliações dos especialistas, apesar da estrutura padronizada da escala Likert, as respostas podem variar conforme a experiência individual dos cirurgiões. O uso da IA na prática médica deve ser realizada de forma criteriosa e responsável, assegurando que sua implementação reforce e amplie o raciocínio clínico, sem comprometer o julgamento médico especializado. A incorporação responsável da IA na prática médica exige rigor científico e um olhar atento às suas limitações, assegurando que seu uso seja um suporte, e não um substituto, para a expertise dos profissionais de saúde.

CONCLUSÃO

A análise quantitativa revelou que as respostas do ChatGPT 4 foram classificadas como “totalmente adequadas” em pouco mais da metade das avaliações, sugerindo que, embora a IA possua um potencial promissor para auxiliar na tomada de decisão clínica, ainda necessita de refinamentos para atingir um nível de confiabilidade compatível com a prática médica. A maior disparidade foi observada na seleção da técnica cirúrgica, onde a IA demonstrou dificuldades em captar nuances individuais dos pacientes e especificidades do procedimento.

REFERÊNCIAS

1. ALI A, et al. Performance of ChatGPT and GPT-4 on Neurosurgery Written Board Examinations, *Neurosurgery*, 2023; 9(6): 1353–1365.
2. ASKARI A, et al. The surgical management of obesity. *Clinical Medicine*, 2023; 23(4): 330–336.
3. CHIESA-ESTOMBA CM, et al. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. *European archives of oto-rhino-laryngology: official journal of the European Federation of Otorhinolaryngological Societies (EUFOS): affiliated with the German Society for Otorhinolaryngology - Head and Neck Surgery*, 2023; 281(4): 2081-2086.
4. COURCOULAS AP, et al. Long term outcomes of metabolic/bariatric surgery in adults. *BMJ*, 2023; 383: 71027.
5. GUMMADI R, et al. Evaluating the accuracy of large language model (ChatGPT) in providing information on metastatic breast cancer, *Advanced Pharmaceutical Bulletin*, 2024; 14(3): 499–503.
6. JAZI AHD, et al. Bariatric Evaluation Through AI: a Survey of Expert Opinions Versus ChatGPT-4 (BETA-SEO), *Obesity Surgery*, 2023; 33(12): 3971–3980.
7. LEE Y, et al. Harnessing Artificial Intelligence in Bariatric Surgery: Comparative Analysis of ChatGPT-4, Bing, and Bard in Generating Clinician-Level Bariatric Surgery Recommendations, *Surgery for obesity and related diseases*, 2024; 20(7): 603-608.

8. LONG C, et al. Evaluating ChatGPT-4 in otolaryngology–head and neck surgery board examination using the CVSA model. *MedRxiv*, 2023; 23290758.
9. LOPES D. The Role of Fundoplication Sleeve Gastrectomy in Obesity Surgery: A Systematic Review and Meta-Analysis. *Repositorio-aberto.up.pt*, 2024.
10. LOPEZ-GONZALEZ R, et al. Evaluation of the Impact of ChatGPT on the Selection of Surgical Technique in Bariatric Surgery. *Obesity Surgery*, 2024.
11. MUSTAFA A, et al. How does ChatGPT perform on the European Board of Pediatric Surgery examination? A randomized comparative study. *Research Square (Research Square)*, 2023.
12. NUNES BC, et al. Técnicas Bariátricas: Vantagens, desvantagens e indicações das principais técnicas cirúrgicas. *Brazilian Journal of Implantology and Health Sciences*, 2024; 6(10): 1737–1749.
13. OH N, et al. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. *Annals of Surgical Treatment and Research*, 2023; 104(5): 269–273.
14. OZSOY Z e DEMIR E. Which Bariatric Procedure Is the Most Popular in the World? A Bibliometric Comparison. *Obesity Surgery*, 2018; 28(8): 2339–2352.
15. PANTELIS A. Current and Potential Applications of Artificial Intelligence in Metabolic Bariatric Surgery. *Bariatric Surgery - Past and Present*, 2023.
16. PAULO D, et al. Análise dos fatores que levam mulheres médicas a não optarem por especialidades cirúrgicas. *Revista de Medicina*, 2020; 99(3): 230–235.
17. PERDOMO CM, et al. Contemporary medical, device, and surgical therapies for obesity in adults. *The Lancet*, 2023; 401(10382): 1116-1130.
18. PREMKUMAR A, et al. Factors Associated With Bariatric Surgery Referral Patterns: A Systematic Review. *Journal of Surgical Research*, 2022; 276: 54–75.
19. RAJEEV ND, et al. Patient and the Public's Perceptions of Bariatric Surgery: A Systematic Review. *Journal of Surgical Research*, 2023; 283: 385–406.
20. SCHEFFER M, et al. Demografia Médica no Brasil 2023. São Paulo, SP: FMUSP, AMB, 2023; 344.
21. SRINIVASAN N, et al. Large language models and bariatric surgery patient education: a comparative readability analysis of GPT-3.5, GPT-4, Bard, and online institutional resources. *Surgical Endoscopy*, 2024; 38(5): 2522–2532.
22. TARIQ R e KHANNA S. Artificial Intelligence–Powered Patient Education for Comprehensive and Individualized Understanding for Patients. *Clinical Gastroenterology and Hepatology*, 2023; 22(7): 1550-1551.
23. TEMSAH O, et al. Overview of Early ChatGPT's Presence in Medical Literature: Insights From a Hybrid Literature Review by ChatGPT and Human Experts. *Cureus*, 2023; 15(4): 37281.