



CENTRO UNIVERSITÁRIO GOYAZES – UNIGOYAZES
CURSO DE ODONTOLOGIA

**CONFIABILIDADE E COMPLETUDE DAS RESPOSTAS DE INTELIGÊNCIAS
ARTIFICIAIS GENERATIVAS (CHATGPT E GEMINI) EM ODONTOPEDIATRIA:
COMPARAÇÃO COM EVIDÊNCIAS CIENTÍFICAS DA PUBMED – ESTUDO
PILOTO**

Maria Eduarda Silva Jesus

Marcos Vinícius Pedroso Rosa

Orientador: Prof. Dr. Arthur Wilson Florencio Costa

Trindade-GO

2025

CENTRO UNIVERSITÁRIO GOYAZES – UNIGOYAZES
CURSO DE ODONTOLOGIA

**CONFIABILIDADE E COMPLETUDE DAS RESPOSTAS DE INTELIGÊNCIAS
ARTIFICIAIS GENERATIVAS (CHATGPT E GEMINI) EM ODONTOPEDIATRIA:
COMPARAÇÃO COM EVIDÊNCIAS CIENTÍFICAS DA PUBMED –
ESTUDO PILOTO**

Maria Eduarda Silva Jesus

Marcos Vinícius Pedroso Rosa

Trabalho de Conclusão de Curso
apresentado ao Centro Universitário
GOYAZES – UniGoyazes como requisito
parcial à obtenção do título de Bacharel
em Odontologia.

Orientador: Prof. Dr. Arthur Wilson Florencio Costa

**Trindade - GO
2025**

Maria Eduarda Silva Jesus
Marcos Vinícius Pedroso Rosa

**CONFIABILIDADE E COMPLETUDE DAS RESPOSTAS DE INTELIGÊNCIAS
ARTIFICIAIS GENERATIVAS (CHATGPT E GEMINI) EM ODONTOLOGIA:
COMPARAÇÃO COM EVIDÊNCIAS CIENTÍFICAS DA PUBMED –
ESTUDO PILOTO**

Trabalho de Conclusão de Curso
apresentado ao Centro Universitário
Goyazes – UniGoyazes como requisito
parcial à obtenção do título de Bacharel em
Odontologia, aprovada pela seguinte
banca examinadora:

Centro
Goyazes

Documento assinado digitalmente
ANNA ALICE ANABUKI
Data: 13/12/2025 08:56:20-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. Arthur Wilson Florencio da Costa (Orientador)

Universitário União de
- Unigoyazes

govbr Documento assinado digitalmente
ARTHUR WILSON FLORENCIO DA COSTA
Data: 13/12/2025 07:33:01-0300
Verifique em <https://validar.itl.gov.br>

Profa. Dra. Anna Alice Anabuki (Membro interno)
Centro Universitário União de Goyazes - Unigoyazes

govbr Documento assinado digitalmente
VICTOR HUGO DE SOUZA UTIDA
Data: 12/12/2025 22:44:42-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. Victor Hugo de Souza Utida (Membro externo)
Pontifícia Universidade Católica



Profa. Esp. Débora Lacerda (Suplente)
Centro Universitário União de Goyazes - UniGOYAZES

**Trindade - GO
2025**

**CONFIABILIDADE E COMPLETUDE DAS RESPOSTAS DE INTELIGÊNCIAS
ARTIFICIAIS GENERATIVAS (CHATGPT E GEMINI) EM ODONTOPEDIATRIA:
COMPARAÇÃO COM EVIDÊNCIAS CIENTÍFICAS DA PUBMED –
ESTUDO PILOTO**

Maria Eduarda S. Jesus¹; Marcos Vinícius P. Rosa¹; Arthur Wilson F. da Costa².

RESUMO

INTRODUÇÃO: O avanço das tecnologias digitais, especialmente da Inteligência Artificial (IA), tem impactado a Odontologia. Ferramentas como ChatGPT e GEMINI são amplamente utilizadas por estudantes e profissionais para estudo e suporte clínico. Entretanto, ainda são escassas as pesquisas que avaliam se suas respostas estão alinhadas com a literatura científica atual. **OBJETIVOS:** Analisar se as respostas do ChatGPT (versões paga e gratuita) e do Gemini (pago e gratuito) sobre temas em Odontopediatria e o quanto estes correspondem aos conhecimentos presentes em revisões sistemáticas da PubMed. **METODOLOGIA:** Este estudo transversal, descritivo e analítico que avaliou a confiabilidade e a completude das respostas fornecidas por quatro plataformas de inteligência artificial (ChatGPT gratuito e pago e Gemini gratuito e pago) em comparação com evidências científicas de alto nível. Cinco perguntas clínicas relevantes em Odontopediatria foram selecionadas e respondidas utilizando revisões sistemáticas e guidelines identificados por meio de busca estruturada na base PubMed, constituindo o padrão-ouro. As mesmas perguntas foram submetidas, de forma padronizada às plataformas de IA. As respostas foram analisadas quanto à confiabilidade e completude. **RESULTADOS:** As quatro plataformas de IA apresentaram desempenho variável. A versão paga do ChatGPT demonstrou a maior confiabilidade, completude das respostas, precisão conceitual e melhor alinhamento com as evidências científicas quando comparada à versão paga do Gemini e às versões gratuitas. Já as versões gratuitas apresentaram maior tendência a respostas genéricas, superficiais e com menor rigor técnico, insuficientes para aplicações acadêmicas ou clínicas que demandem precisão metodológica. **CONSIDERAÇÕES FINAIS:** Embora o ChatGPT pago tenha se destacado em relação aos demais modelos, nenhuma das IAs avaliadas alcançou consistência suficiente para substituir revisões sistemáticas ou guias clínicos como fontes primárias de informação científica.

PALAVRAS-CHAVE: Odontopediatria, Inteligência Artificial, Confiabilidade.

¹Acadêmico do Curso de ODONTOLOGIA do Centro Universitário Goyazes – UniGoyazes

²Orientador: Prof. Doutor, do Centro Universitário Goyazes – UniGoyazes.

RELIABILITY AND COMPLETENESS OF RESPONSES FROM GENERATIVE ARTIFICIAL INTELLIGENCES (CHATGPT AND GEMINI) IN PEDIATRIC DENTISTRY: COMPARISON WITH SCIENTIFIC EVIDENCE FROM PUBMED – PILOT STUDY

Maria Eduarda S. Jesus¹; Marcos Vinícius P. Rosa¹; Arthur Wilson F. da Costa².

ABSTRACT

INTRODUCTION: The advancement of digital technologies, especially Artificial Intelligence (AI), has impacted Dentistry. Tools such as ChatGPT and GEMINI are widely used by students and professionals for study and clinical support. However, research evaluating whether their answers align with current scientific literature is still scarce. **OBJECTIVES:** To analyze whether the answers from ChatGPT (paid and free versions) and Gemini (paid and free) on topics in Pediatric Dentistry correspond to the knowledge present in systematic reviews from PubMed. **METHODOLOGY:** This cross sectional, descriptive, and analytical study evaluated the reliability and completeness of the answers provided by four artificial intelligence platforms (free ChatGPT, paid ChatGPT, free Gemini, and paid Gemini) in comparison with high-level scientific evidence. Five relevant clinical questions in Pediatric Dentistry were selected and answered using systematic reviews and guidelines identified through a structured search in the PubMed database, constituting the gold standard. The same questions were submitted, in a standardized way, to the AI platforms. The responses were analyzed for reliability and completeness. **RESULTS:** The four AI platforms showed varying performance. The paid version of ChatGPT demonstrated the highest reliability, completeness of answers, conceptual accuracy, and best alignment with scientific evidence when compared to the paid version of Gemini and the free versions. The free versions, however, showed a greater tendency towards generic, superficial answers with less technical rigor, insufficient for academic or clinical applications that demand methodological precision. **FINAL CONSIDERATIONS:** Although the paid version of ChatGPT stood out compared to the other models, none of the AIs evaluated achieved sufficient consistency to replace systematic reviews or clinical guidelines as primary sources of scientific information.

KEYWORDS: Pediatric Dentistry, Artificial Intelligence, Reliability.

¹Student of the Dentistry Course at the Goyazes University Center – UniGoyazes.

²Advisor: Professor Doctor, from the Goyazes University Center – UniGoyazes.

AGRADECIMENTOS

Agradecemos primeiramente a Deus, nosso refúgio e fortaleza durante toda a caminhada. Foi Ele quem nos sustentou nos momentos de cansaço, nos deu esperança nos dias de incerteza e sabedoria para seguir em frente. Reconhecemos que, sem Sua presença constante, este trabalho não teria sido possível.

Ao nosso orientador, Prof. Arthur Wilson, expressamos nossa profunda gratidão pela dedicação, paciência e generosidade ao compartilhar seu conhecimento. Sua orientação foi essencial para a construção deste trabalho, não apenas pelo apoio científico, mas também pelo incentivo e pela confiança depositada em nós.

Aos nossos pais, agradecemos por todo amor, força e apoio incondicional. Foram eles que acreditaram em nós mesmo quando duvidávamos de nós mesmos, ofereceram segurança nos tropeços e celebraram cada conquista ao nosso lado. Este trabalho é fruto do esforço, do cuidado e da fé de vocês. “Dedico todo e qualquer sucesso aos meus pais, que, sob muito sol, me fizeram chegar aqui pela sombra e água fresca.”

Agradecemos ainda aos familiares, amigos, professores e colegas que, de diferentes formas, contribuíram para esta jornada. Cada palavra de incentivo, cada gesto de carinho e cada troca de conhecimento foram fundamentais para que chegássemos até aqui. Levaremos para sempre a memória e a gratidão por tudo o que fizeram por nós.

“Não fui eu que lhe ordenei? Seja forte e corajoso! Não se apavore nem se desanime, pois o Senhor, o seu Deus, estará com você por onde você andar.”

(Josué 1:9)

SUMÁRIO

1. INTRODUÇÃO.....	09
JUSTIFICATIVA.....	11
OBJETIVOS.....	12
3.1OBJETIVO GERAL.....	12
OBJETIVOS ESPECÍFICOS.....	12
METODOLOGIA.....	13
DELINEAMENTO DA PESQUISA.....	13
4.2 FORMULAÇÃO E SELEÇÃO DAS PERGUNTAS E PESQUISA.....	13
4.2.1. CRITÉRIOS PARA SELEÇÃO DAS PERGUNTAS.....	13
4.2.2. FORMULAÇÃO FINAL DAS PERGUNTAS.....	13
4.3 IDENTIFICAÇÃO DO PADRÃO-OURO.....	14
4.3.1 JUSTIFICATIVA PARA ESCOLHA LITERATURA CIENTÍFICA	14
4.4 BASE DE DADOS CONSULTADA.....	14
4.5 ESTRATÉGIA DE BUSCA NA LITERATURA.....	14
4.5.1 PROCEDIMENTOS DE BUSCA.....	14
4.5.2 PERÍODO DE PUBLICAÇÃO.....	14
4.6 CRITÉRIOS DE INCLUSÃO.....	14
4.7 CRITÉRIOS DE EXCLUSÃO.....	15
4.8 EXTRAÇÃO E SÍNTESE DAS EVIDÊNCIAS CIENTÍFICAS.....	15
4.8.1 PROCEDIMENTO DE EXTRAÇÃO DE DADOS.....	15
4.9 ELABORAÇÃO DAS RESPOSTAS PADRÃO-OURO.....	15

4.10 VALIDAÇÃO DO PADRÃO-OURO.....	16
4.11 SELEÇÃO DAS PLATAFORMAS DE INTELIGÊNCIA ARTIFICIAL.....	16
4.12 COLETA DAS RESPOSTAS DAS IAS.....	16
4.12.1 PROCEDIMENTO PADRONIZADO.....	16
4.13 CONTROLE DE VARIÁVEIS	17
4.14 CRITÉRIOS DE AVALIAÇÃO DAS RESPOSTAS	17
4.15 AVALIAÇÃO DA CONFIABILIDADE	17
4.16 AVALIAÇÃO DA COMPLETUDE	17
4.16.1 CÁLCULO	17
4.17 CÁLCULO FINAL DA PONTUAÇÃO	18
4.18 ANÁLISE DOS DADOS	18
4.18.1 ORGANIZAÇÃO ESTATÍSTICA	18
4.19 JUSTIFICATIVA PARA AUSÊNCIA DE TESTES INFERÊNCIAS	18
4.20 GARANTIA DE RIGOR METODOLÓGICO	18
4.21 CONSIDERAÇÕES ÉTICAS	19
4.22 LIMITAÇÕES METODOLÓGICAS	19
5. RESULTADOS.....	20
6. DISCUSSÃO.....	27
7. CONSIDERAÇÕES FINAIS.....	30
8. REFERÊNCIAS	31

1. INTRODUÇÃO

A tecnologia tem avançado muito nas últimas décadas, especialmente com o surgimento da Inteligência Artificial (IA) (Savegnago et al., 2024). Atualmente, duas das mais utilizadas são o ChatGPT e o Gemini. Esses modelos de IA, desenvolvidos pela OpenAI e pelo Google, são ferramentas capazes de produzir textos e responder perguntas de forma rápida e aparentemente segura, o que tem despertado interesse em diversas áreas do conhecimento, inclusive na Odontologia (NAEIMI *et al.*, 2024).

Na Odontologia — e, mais especificamente, na Odontopediatria — o uso de tecnologias como o ChatGPT e o Gemini está se tornando cada vez mais comum

entre estudantes, professores e até mesmo profissionais que buscam informações rápidas sobre temas específicos. Por serem ferramentas úteis para esclarecer dúvidas, revisar conteúdos e até auxiliar na prática clínica, é fundamental verificar se as informações fornecidas são realmente corretas e atualizadas, já que essas tecnologias não substituem o conhecimento científico baseado em evidências (VISHWANATHAIAH *et al.*, 2023; MAHAJAN *et al.*, 2023).

A Odontopediatria é uma especialidade que exige grande responsabilidade, pois envolve o atendimento de crianças e adolescentes em diferentes fases do desenvolvimento. Tomar decisões clínicas baseadas em informações incorretas pode prejudicar o tratamento e comprometer a saúde dos pacientes. Por isso, é essencial garantir que qualquer fonte de informação — incluindo a IA — seja confiável (ALHARBI & ALHARBI, 2024).

Ainda existem poucas pesquisas que analisam o conteúdo gerado por esse tipo de ferramenta e verificam se ele está realmente de acordo com a literatura científica. Isso torna necessário investigar com mais profundidade se as respostas fornecidas pela IA condizem com os conhecimentos mais atualizados. Uma forma eficaz de realizar essa verificação é comparar as informações fornecidas pela IA com artigos científicos publicados em revistas reconhecidas, especialmente aqueles que seguem a metodologia de revisão sistemática, considerada uma das mais confiáveis dentro da pesquisa científica (DINÇ, KOCAAYDIN & ILLISULU, 2025; GULSEN & GULSEN, 2025).

Neste trabalho, essa comparação será realizada por meio de perguntas previamente elaboradas e apresentadas ao ChatGPT e ao Gemini, abordando temas importantes da Odontopediatria. Em seguida, o conteúdo das respostas será confrontado com estudos do tipo revisão sistemática com metanálise e guidelines encontrados na base de dados PubMed (DINÇ, KOCAAYDIN & ILLISULU, 2025). O objetivo é avaliar até que ponto a IA pode ser considerada uma fonte segura de informações na área odontopediátrica.

2. JUSTIFICATIVA

O uso de ferramentas de inteligência artificial, como o ChatGPT e o Gemini, tem crescido significativamente entre estudantes e profissionais da área da saúde. Essas tecnologias são frequentemente utilizadas para estudos, esclarecimento de dúvidas e até para apoio na preparação de atendimentos clínicos. Entretanto, nem sempre é possível garantir que as informações fornecidas por esses sistemas sejam corretas, atualizadas ou fundamentadas em evidências científicas. Na

Odontopediatria, essa preocupação assume ainda maior relevância, pois a prática clínica envolve o cuidado com crianças e adolescentes, populações que demandam condutas precisas e seguras. Informações imprecisas podem resultar em erros diagnósticos ou terapêuticos, comprometendo a saúde do paciente. Diante disso, torna-se essencial avaliar se os conteúdos gerados por inteligência artificial estão alinhados com o conhecimento científico disponível na literatura especializada, garantindo que seu uso seja responsável e não substitua a tomada de decisão baseada em evidências.

3. OBJETIVOS

3.1 Objetivo geral

Avaliar o quão as respostas fornecidas pelas IAs (Chat GPT pago, Chat GPT gratuito, Gemini pago e Gemini gratuito) sobre questionamentos relacionados a prática clínica da Odontopediatria estão de acordo com os conhecimentos científicos disponíveis em revisões sistemáticas recentes publicadas na base de dados

PubMed.

3.2Objetivos específicos

- Comparar a acurácia das respostas das IAs com as melhores evidências científicas disponíveis (PubMed) em Odontopediatria.
- Classificar o nível de concordância entre as respostas das IAs e as revisões sistemáticas/guidelines da PubMed.
- Identificar inconsistências, desatualizações ou potenciais riscos de desinformação presentes nas respostas.
- Mapear os temas da Odontopediatria que apresentam maior divergência entre a IA e a literatura científica atual.

4. METODOLOGIA

4.1Delineamento da pesquisa

Este estudo caracteriza-se como um estudo clinimétrico com abordagem, cujo objetivo foi avaliar a confiabilidade e a completude das respostas fornecidas por diferentes plataformas de Inteligência Artificial (IA) em comparação com evidências

científicas de alta qualidade provenientes de revisões sistemáticas e guidelines clínicos relacionados à Odontopediatria.

4.2 Formulação e seleção das perguntas de pesquisa

4.2.1 Critérios para seleção das perguntas

As cinco perguntas avaliadas neste estudo foram previamente selecionadas pelo orientador (especialista, mestre e doutor em Odontopediatria) mediante os seguintes critérios:

- Alta relevância clínica;
- Impacto direto na tomada de decisões em Odontopediatria;
- Existência de diretrizes e revisões sistemáticas recentes sobre os temas;
- Possibilidade de comparar respostas de forma objetiva (confiabilidade e completude);
- Recorrência dessas dúvidas em contextos acadêmicos e assistenciais.

4.2.2 Formulação final das perguntas

As perguntas contemplaram temas relacionados a:

- Terapêutica medicamentosa em Odontopediatria;
- Condutas frente a cenários clínicos específicos;
- Tratamento restaurador e pulpar baseado em evidências;
- Manejo de urgências odontopediátricas;
- Recomendações atualizadas da literatura científica.

4.3 Identificação do padrão-ouro (gold standard)

4.3.1 Justificativa para escolha da literatura científica

Para definir o padrão-ouro utilizado como referência para comparação, optou-se exclusivamente por revisões sistemáticas com ou sem meta-análise, guidelines clínicos de sociedades científicas e instituições reconhecidas.

A escolha se baseia no fato de que esses estudos representam os mais altos níveis de evidência científica disponíveis, reduzindo vieses e aumentando a validade das recomendações.

4.4 Base de dados consultada

A busca foi realizada na base PubMed/MEDLINE. Essa base foi escolhida por abranger a maior quantidade de estudos biomédicos revisados por pares, garantindo qualidade e rigor científico.

4.5 Estratégia de busca na literatura

4.5.1 Procedimentos de busca

A busca foi feita manualmente, utilizando descritores MeSH e termos livres combinados com operadores booleanos (AND/OR/NOT). Cada pergunta recebeu uma estratégia de busca própria (ex.: “pulpotomy”, “pit and fissure sealants”, “fluoride therapy”, “trauma dental – guidelines”, etc.).

4.5.2 Período de publicação

Foram incluídos artigos e diretrizes publicados nos últimos 10 anos, com o intuito de garantir atualidade das recomendações utilizadas como padrão-ouro.

4.6 Critérios de inclusão

Foram considerados como critérios de inclusão:

- Revisões sistemáticas com ou sem meta-análise.
- Guidelines clínicos oficiais.

14

- Estudos em inglês, português ou espanhol.
- Estudos diretamente relacionados ao conteúdo das perguntas.

4.7 Critérios de exclusão

Foram considerados como critérios de exclusão:

- Revisões narrativas, cartas, comentários e opiniões de especialistas.
- Estudos que não abordavam o foco da pergunta.
- Artigos duplicados.

4.8 Extração e síntese das evidências científicas

4.8.1 Procedimento de extração de dados

Para cada revisão sistemática e guideline selecionado foram extraídos:

- Recomendações clínicas;
- Opções terapêuticas;
- Doses, intervalos e medicamentos (quando aplicável);
- Critérios diagnósticos;
- Evidências de eficácia;
- Grau de recomendação (se informado).

4.9 Elaboração das respostas padrão-ouro

As informações extraídas foram sintetizadas e organizadas em respostas objetivas e completas para cada uma das cinco perguntas. Cada resposta padrão ouro foi estruturada contendo:

- Todos os itens essenciais recomendados;
- Indicações e contraindicações;
- Condutas preferenciais segundo a literatura;
- Pontos críticos e fundamentos clínicos;
- Justificativas quando presentes na revisão sistemática/meta-análise.

15

Essas respostas foram utilizadas como base de comparação para avaliação das Inteligências Artificiais.

4.10 Validação do padrão-ouro

Dois avaliadores previamente treinados obtiveram as respostas eleitas como

padrão-ouro com base na evidência científica e um terceiro avaliador especialista na área, revisou e validou as respostas padrão-ouro para garantir os aspectos de confiabilidade, completude e o alinhamento estrito com as fontes científicas selecionadas.

4.11 Seleção das plataformas de Inteligência Artificial

Foram avaliadas quatro plataformas: ChatGPT Gratuito (OpenAI), ChatGPT Pago (OpenAI), Gemini Gratuito (Google) e Gemini Pago (Google). Essa seleção se justifica por serem as plataformas mais utilizadas no momento e por ofertarem versões distintas de acesso (gratuita e paga), permitindo comparação entre elas.

4.12 Coleta das respostas das IAs

4.12.1 Procedimento padronizado de coleta

Para evitar vieses e interferências, foi adotado o seguinte protocolo:

- Cada IA foi acessada em um dispositivo diferente, sem histórico prévio. •
 - Não foram fornecidos prompts adicionais além da pergunta.
 - A pergunta foi copiada e colada integralmente.
 - As respostas foram registradas tabeladas e resumidas.
 - Cada resposta foi identificada pela plataforma correspondente. •
- Todas as respostas foram transcritas para uma planilha de análise.

16

4.13 Controle de variáveis

As perguntas foram enviadas no mesmo dia para todas as plataformas. O horário de envio foi padronizado para evitar instabilidade dos servidores. Nenhuma resposta foi complementada com perguntas adicionais.

4.14 Critérios de avaliação das respostas das IAs

A análise consistiu em dois parâmetros principais: Confiabilidade (0–5 pontos) e Completude (0–5 pontos) das respostas fornecidas pelas inteligências artificiais.

4.15 Avaliação da confiabilidade

A confiabilidade das respostas fornecidas pelas inteligências artificiais foi avaliada conforme apresentado abaixo:

Nível	Critério	Pontuação
Confiável	A resposta está totalmente de acordo com o padrão-ouro, sem erros científicos.	5 pontos
Parcialmente confiável	Contém informações corretas, mas apresenta pequenas falhas, omissões ou inconsistências.	2,5 pontos
Não confiável	Contém erros importantes, condutas incorretas ou desatualizadas.	0 ponto

A análise da confiabilidade foi realizada pelos pesquisadores identificando a qual IA pertencia cada resposta.

4.16 Avaliação da completude

A completude das respostas foi avaliada com base na quantidade de itens essenciais presentes na resposta padrão-ouro.

4.16.1 Cálculo

Se a resposta padrão-ouro continua X itens então o valor total dos X itens foi igual a 5,0 pontos e o valor por item foi considerado dividindo os 5,0 pontos da

completude pelo total de itens. Sendo assim, as respostas das IAs recebiam

pontuação proporcional ao número de itens fornecidos corretamente.

Exemplos:

- 2 itens fornecidos → 5 pontos (cada item vale 2,5);
- 1 item fornecidos → Cada item 2,5, logo apenas 1 item fornecido é igual 2,5. •
- 0 itens fornecidos → 0 pontos.

Para os casos em que a confiabilidade, avaliada anteriormente ao item de completude, resultava em nota zero automaticamente o item de completude pontuava também como zero, uma vez que o item não era confiável.

4.17 Cálculo final da pontuação

Inicialmente cada resposta recebeu uma nota de confiabilidade (0–5) e uma nota de completude (0–5) e a soma dos dois itens resultou em uma nota final (0-10) para cada pergunta. Em seguida a média final de confiabilidade e completude juntos correspondeu a somatória das medias de cada pergunta (5 perguntas) dividido por 5.

4.18 Análise dos dados

4.18.1 Organização estatística

Foi realizada uma estatística descritiva e numérica das respostas obtidas e estas foram analisadas por dois avaliadores calibrados.

4.19 Justificativa para ausência de testes inferenciais

A pesquisa não visa estabelecer inferência populacional, mas sim comparar desempenho entre plataformas de inteligência artificial específicas. Assim, métodos descritivos atendem aos objetivos do estudo.

4.20 Garantia de rigor metodológico

Para assegurar validade e reprodutibilidade do estudo foram adotadas as seguintes medidas:

- Padronização do não uso de prompts;

- Coleta simultânea das respostas;
- Avaliação cega das respostas pelos examinadores;
- Definição clara de critérios de pontuação;
- Uso de literatura científica de alto nível como padrão-ouro.

4.21 Considerações éticas

Por não envolver seres humanos nem dados sensíveis, o estudo está dispensado de submissão ao Comitê de Ética em Pesquisa, conforme Resolução CNS nº 510/2016.

4.22 Limitações metodológicas

As limitações desse estudo contemplam aspectos como: (1) As inteligências artificiais são atualizadas continuamente, podendo modificar seu desempenho ao longo do tempo. (2) A interpretação da completude exige julgamento técnico, apesar dos critérios objetivos. (3) O estudo avaliou somente 5 perguntas, o que pode não representar todo o campo da Odontopediatria. (4) A coleta foi feita em um único momento, sem avaliação longitudinal. (5) A análise é dependente da clareza e especificidade das respostas das IAs.

5. RESULTADOS

Os resultados mostram que as revisões sistemáticas e guidelines podem ser considerados como padrão ouro na qualidade e completude de informações em relação aos temas pesquisados. Em comparação a isso, a inteligência artificial Chat GPT pago foi mais completa e confiável dentre todas as IAs, seguido do Gemini pago, Chat GPT gratuito e Gemini gratuito, respectivamente.

Na avaliação da pergunta de número 1 foi observado que o pior resultado foi disponibilizado pelo Chat GPT gratuito (2,5) seguido do Gemini gratuito (7,5) e o Gemini pago (8,75). Já a IA Chat GPT pago (10,0) mostrou um nível de evidência forte e completo para a resposta da pergunta quando comparado com a evidência fornecida pela base de dados PubMed (10,0).

Foi notado durante a avaliação das respostas fornecidas para a pergunta 2 que o pior resultado foi do Gemini gratuito (2,5) seguido do Chat GPT Gratuito (5,82) e Gemini pago (5,82). A IA Chat GPT pago apresentou uma nota melhor (10,0) assim como o PubMed que obtiveram notas máximas (10,0).

Durante a avaliação da pergunta 3 as inteligências artificiais Gemini gratuito (2,5), Gemini pago (2,5) e Chat GPT gratuito (2,5) obtiveram as piores notas, enquanto o Chat GPT pago (10,0) e o PubMed se equipararam com as notas máximas (10,0).

Ao analisarmos às respostas para a pergunta 4 foi observado que o pior resultado obtido foi pela IA Gemini pago (2,5). Porém todas as outras respostas fornecidas pelas demais IAs mostraram um nível de evidência forte e completo da resposta da pergunta quando comparado com as evidências fornecidas pela base de dados do PubMed (10,0).

A avaliação das respostas para a pergunta 5 mostrou que o Gemini gratuito (2,5) e Gemini pago (2,5) forneceu o pior resultado, seguido do ChatGPT gratuito (4,16) e Chat GPT pago (5,82). Sendo assim nenhuma das IA conseguiu apresentar nível de confiabilidade e completude igual ao PubMed (10,0).

O detalhes dos resultados encontrados nesse estudo podem ser observados detalhadamente no quadro de número 1.

Quadro 1. Compilado das respostas fornecidas pelas revisões sistemática e pelas inteligências artificiais generativas frente as perguntas do estudo.

Pergunta	Revisão sistemática PUBMED	ChatGPT G	ChatGPT P	Gemini
Pergunta 1: Qual antibiótico prescrever em crianças que são alérgicas à amoxicilina na odontopediatria?	A International Association of Paediatric Dentistry, em suas recomendações, afirma que em caso de alergia a penicilinas, pode se usar: Clindamicina, azitromicina, cefalosporina ou claritromicina. <i>International Association of Paediatric Dentistry. Use of Antibiotic Therapy for Paediatric Dentistry (2021).</i>	Quando a criança é alérgica à amoxicilina (ou à penicilina em geral), é preciso escolher um antibiótico de outra classe para evitar reações cruzadas. A escolha depende do tipo de infecção e da gravidade do quadro.	Opções de 1ª linha: Azitromicina, claritromicina e clindamicina. Se a alergia NÃO for do tipo imediato (sem anafilaxia/urticária): Cefalexina - Cefalosporina.	Para crianças com alergia à amoxicilina: 1. Macrolídeos: Azitromicina, claritromicina e Eritromicina. 2. Lincosamidas: Clindamicina. 3. Cefalosporinas: Cefalexina.
Confiabilidade	5,0	2,5	5,0	2,5
Completude	5,0	0,0	5,0	5,0
Nota final	10,0	2,5	10,0	7,5

Pergunta	Revisão sistemática PUBMED	ChatGPT G	ChatGPT P	Gemini
----------	----------------------------	-----------	-----------	--------

Pergunta 2: Qual a melhor pasta obturadora para dentes decíduos tratados endodonticamente?	Não existe uma única pasta obturadora de dentes decíduos. As opções são: Pasta de hidróxido de cálcio + iodoformio (Ca (OH)₂/iodoform) – Metapex, Vitapex e Feapex. Pasta ZOE (Óxido de Zinco com Eugenol). Pasta ZOE + Iodofórmio combinado com Ca (OH)₂ - Endoflas. <i>Najjar et al., 2019 – Revisão sistemática com metanálise.</i>	Não há uma “melhor” pasta obturadora para dentes decíduos tratados endodonticamente que seja consensual, mas há sim materiais com boas evidências e perfis distintos. Pasta ZOE (Óxido de Zinco e Eugenol). Pastas de iodofórmio + Ca (OH)₂ – Vitapex, Metapex.	Não existe “a melhor” única – as diretrizes e revisões mostram boas taxas de sucesso com mais de um material. Opções com melhor evidência atual: Pasta de hidróxido de cálcio + Iodofórmio – Vitapex, Metapex. Pasta ZOE + Ca (OH)₂ + Iodofórmio – Endoflas. Pasta ZOE (Óxido de Zinco com Eugenol).	A escolha da melhor pasta obturadora para dentes decíduos tratados endodonticamente (pulpa vital) é um tema que não possui consenso na literatura odontológica.
Confiabilidade	5,0	2,5	5,0	2,5
Completude	5,0	3,32	5,0	0,0
Nota final	10,0	5,82	10,0	2,5

22

Pergunta	Revisão sistemática PUBMED	ChatGPT G	ChatGPT P	Gemini
Pergunta 3: O uso de anestésico tópico pode ser usado para aliviar o desconforto do nascimento dos dentes em bebês?	Não usar produtos orais que contenham benzocaína para aliviar dor de erupção dentária em bebês e crianças menores de 2 anos — risco de metemoglobinemia, potencialmente fatal. Não usar solução viscosa de lidocaína 2% para tratar dor de dentição em lactentes.	Não é recomendado o uso de anestésicos tópicos (géis, pomadas ou sprays) para aliviar a dor da erupção dentária em bebês.	Não, o uso de anestésicos tópicos (como pomadas com lidocaína e benzocaína) não é recomendada para aliviar o desconforto do nascimento dos dentes em bebês.	O uso de anestésicos tópicos (géis ou pomadas contendo lidocaína e benzocaína) para aliviar o desconforto no nascimento dos dentes em bebês é geralmente recomendado e deve ser usado com extrema cautela, sob supervisão médica, e seguindo as orientações da literatura.

	<i>FDA – Food and Drug Administration (2018).</i>			profissionais de saúde.
Confiabilidade	5,0	2,5	5,0	2,5
Completude	5,0	0,0	5,0	0,0
Nota final	10,0	2,5	10,0	2,5

23

Pergunta	Revisão sistemática PUBMED	ChatGPT G	ChatGPT P	Gemini
Pergunta 4: Qual é o melhor analgésico para tratar dor dentária em crianças?	Recomendação geral: Paracetamol (acetaminofeno) e antiinflamatórios não esteroidais (AINES) – especialmente o ibuprofeno – são as terapias de primeira linha para dor dentária aguda em crianças. <i>Carrasco-Labra et al., 2023 - Revisão sistemática com método GRADE – Guideline.</i>	Primeira escolha: Paracetamol (acetaminofeno). Segunda escolha: Ibuprofeno.	Ibuprofeno (se não houver contraindicação a AINES). Paracetamol (acetaminofeno).	Os medicamentos mais comumente utilizados são considerados eficazes para aliviar a dor dentária em crianças. Paracetamol Ibuprofeno
Confiabilidade	5,0	5,0	5,0	5,0
Completude	5,0	5,0	5,0	5,0
Nota final	10,0	10,0	10,0	10,0

24

Pergunta	Revisão sistemática PUBMED	ChatGPT G	ChatGPT P	Gemini
----------	----------------------------	-----------	-----------	--------

Pergunta 5: Qual é o melhor tratamento para lesões de cárie cavitadas em esmalte de dentes decíduos?	As melhores fontes (diretrizes e revisões sistemáticas) recomendam condutas minimamente invasivas para lesões restritas ao esmalte: Selantes e resina ou ionômero. Resina infiltrante quando tecnicamente possível. Aplicação de verniz fluoretado periódico e monitorização ativa. <i>Slayton et al., (2018) – Revisão sistemática / Guideline.</i>	Opção 1 – Selamento com Selante de resina ou ionômero. Opção 2 – Restauração atraumática (ART). Opção 3 – Diaminofluoreto de prata (DFP /SDF).	Opções de tratamento: Controle de biofilme + Aplicação tópica de flúor a cada 3 meses. Técnica de tratamento restaurador Atraumático (ART). Infiltração resinosa (ICON).	O "melhor" tratamento individualizado, mas as condutas atuais priorizam a máxima preservação da estrutura
Confiabilidade	5,0	2,5	2,5	2,5
Completude	5,0	1,66	3,32	0,83
Nota final	10,0	4,16	5,82	2,33

25

Pergunta	Revisão sistemática PUBMED	ChatGPT G	ChatGPT P	Gemini
Médias finais – Confiabilidade e completude	10,0	4,99	9,16	5,83
Classificação final das fontes	1º lugar	4º lugar	2º lugar	3º lugar

26

6. DISCUSSÃO

Os resultados deste estudo demonstram uma diferença clara entre a qualidade da informação científica acessada por revisões sistemáticas e guidelines — considerados o padrão ouro — e aquela fornecida por ferramentas de inteligência artificial generativa. Essa constatação reforça o entendimento já amplamente discutido na literatura, de que a IA, embora apresente avanços consideráveis e seja capaz de produzir respostas rápidas e coerentes, ainda não alcança o rigor metodológico e a precisão das evidências de alto nível (Ding et al., 2023).

Ao analisar o desempenho consolidado dos modelos, observou-se uma hierarquização marcada: o ChatGPT pago apresentou a maior média (9,16), posicionando-se muito acima das demais ferramentas avaliadas e demonstrando proximidade com o padrão ouro. Já as demais IAs apresentaram um desempenho consideravelmente menor: o ChatGPT gratuito obteve 4,99, seguido do Gemini gratuito (5,0) e do Gemini pago (4,41). Esse agrupamento demonstra que as versões gratuitas — e particularmente o Gemini pago — apresentaram desempenho inferior e relativamente homogêneo, enquanto apenas o ChatGPT pago se destacou de forma expressiva. Esses achados estão de acordo com o estudo de Kunzle e Paris (2024), que relataram desempenho superior e mais consistente em modelos avançados como o ChatGPT.

A análise individual das questões reforça que o desempenho das IAs não é linear e varia substancialmente conforme o conteúdo abordado. Em algumas perguntas, modelos com média inferior, como o Gemini gratuito, apresentaram desempenho relativamente melhor do que versões pagas, indicando que determinados temas ou palavras-chave podem coincidir com pontos fortes específicos no treinamento dos modelos — fenômeno previamente discutido por Buldur e Sezer (2024). Assim, mesmo IAs com desempenho global mais baixo podem, ocasionalmente, produzir respostas de alta qualidade.

A ocorrência de notas muito baixas ou incompatíveis com a literatura para algumas perguntas, especialmente entre os modelos Gemini (gratuito e pago) e o ChatGPT gratuito, confirma achados prévios de Suárez e colaboradores

(2024), que observaram maior frequência de respostas incompletas ou imprecisas em áreas clínicas específicas, como endodontia e dentística restauradora. Isso sugere que determinados campos do conhecimento representam desafios maiores para IAs generativas, possivelmente pela complexidade dos protocolos e pela escassez de exemplos clínicos de alta qualidade no conjunto de treinamento.

A literatura também destaca que o desempenho das IAs pode variar significativamente dentro de uma mesma área temática, o que foi observado neste estudo. O comportamento instável de alguns modelos, alternando entre respostas completas e respostas incompatíveis com a evidência, reforça as conclusões de Mondillo e colaboradores (2024) de que a performance da IA é fortemente dependente da estrutura informacional interna — e não apenas de sua versão (gratuita ou paga).

Embora o ChatGPT pago tenha apresentado desempenho muito superior às demais IAs, ele também demonstrou limitações em perguntas que exigiam maior profundidade conceitual, confirmando o que foi observado no estudo de Water e colaboradores (2023) sobre a tendência dos modelos de IA em fornecer respostas convincentes, porém incompletas ou inconsistentes. Esse achado somado aos nossos resultados reforça que, mesmo em sua versão mais avançada, a IA não substitui o uso de revisões sistemáticas e a interpretação crítica realizada por especialistas.

De modo geral, embora as IAs generativas possam ser ferramentas úteis para estudo, revisão inicial de conteúdos e organização de informações, elas não substituem a consulta às melhores evidências científicas nem o raciocínio clínico fundamentado, como aquele encontrado no PubMed. A considerável distância entre o desempenho do ChatGPT pago e o das demais ferramentas demonstra que a escolha do modelo de IA influencia fortemente a qualidade das respostas obtidas em buscas rápidas, mas ainda assim nenhum deles apresentou desempenho totalmente confiável.

Por fim, este estudo destaca a necessidade de que estudantes e

profissionais da saúde desenvolvam competência crítica para avaliar a qualidade

28

da informação digital, reconhecendo que a IA pode auxiliar na aprendizagem, mas não elimina a responsabilidade do aluno ou profissional em confirmar a confiabilidade e a completude das informações obtidas por meio das IAs.

7. CONSIDERAÇÕES FINAIS

- As revisões sistemáticas e os guidelines permanecem como as fontes

mais confiáveis, completas e seguras para fundamentar decisões clínicas e acadêmicas.

- Embora o ChatGPT em sua versão paga tenha apresentado desempenho significativamente superior aos demais modelos, nenhuma das IAs avaliadas demonstrou consistência suficiente para substituir fontes de evidência científica de alto nível.
- O desempenho das IAs varia de acordo com a plataforma e com a complexidade do tema abordado. Observou-se uma grande oscilação entre as respostas, especialmente nos modelos gratuitos e no Gemini pago, que apresentaram médias baixas.
- Mesmo a versão paga do ChatGPT, apesar de ser a mais próxima do padrão ouro, exibiu limitações importantes em questões que exigiam maior profundidade conceitual.
- As tecnologias de IA generativa devem ser utilizadas exclusivamente como ferramentas complementares de estudo e apoio inicial à busca de informações, nunca como fonte primária para tomada de decisões clínicas ou científicas.
- É importante o uso crítico e responsável dessas ferramentas, especialmente no contexto da saúde, onde precisão e rigor são essenciais.

1. SAVEGNAGO, G.; PINTO, G. V.; SNOVARESKI, C. F.; HAMAD, N. O.; SERPA, G. F.; LIEDKE, G. S. Inteligência artificial na odontologia: uma revisão narrativa de literatura. *Revista da Faculdade de Odontologia – UPF*, v. 29, n. 1, 2024.
2. NAEIMI, S. M.; DARVISH, S.; NAZEMI SALMAN, B.; LUCHIAN, I. Artificial intelligence in adult and pediatric dentistry: a narrative review. *Bioengineering*, v. 11, n. 5, 2024.
3. ALHARBI, N.; ALHARBI, A. S. AI-driven innovations in pediatric dentistry: enhancing care and improving outcome. *Cureus*, 2024.
4. DİNÇ, D.; KOCAAYDIN, S.; İLİSULU, S. C. Are chatbots reliable sources of information regarding fluoride in pediatric dentistry? *Journal of Evidence-Based Dental Practice*, v. 25, n. 1, 2025.
5. GÜLŞEN, E.; GÜLŞEN, İ. T. Utilization of artificial intelligence in pediatric dentistry: a comprehensive literature review. *J Dent Sci Educ.*, v. 3, n. 1, p. 16–19, 2025. DOI: 10.51271/JDSE-0048.
6. MAHAJAN, K.; KUNTE, S. S.; PATIL, K. V.; SHAH, P. P.; SHAH, R. V.; JAJOO, S. S. Artificial intelligence in pediatric dentistry – a systematic review. *Journal of Dental Research and Reviews*, 2023. DOI: 10.4103/jdrr.jdrr_199_22.
7. INTERNATIONAL ASSOCIATION OF PAEDIATRIC DENTISTRY. Use of antibiotic therapy for paediatric dentistry. 2021. Disponível em: https://iapdworld.org/wp-content/uploads/2020/12/01_Use-of-Antibiotic-Therapy-for-Paediatic-Dentistry.pdf. Acesso em: 23 nov. 2025.
8. NAJJAR, RAFAH S.; ALAMOUDI, NAJLAA M.; EL-HOUSSEINY, AZZA A.; AL TUWIRQI, AMANI A.; SABBAGH, HEBA J. A comparison of calcium

hydroxide/iodoform paste and zinc oxide eugenol as root filling materials for pulpectomy in primary teeth: a systematic review and meta-analysis. *Clinical and Experimental Dental Research*, v. 5, n. 4, p. 397–413, 2019. DOI: 10.1002/cre2.173.

9. FOOD AND DRUG ADMINISTRATION (FDA). Letter regarding benzocaine over-the-counter oral health care drug products. Silver Spring, MD: FDA, 21 maio 2018. Disponível em: <https://www.fda.gov/files/drugs/published/FDA-Letter-Regarding-Benzocaine.pdf>. Acesso em: 23 nov. 2025.

10. CARRASCO-LABRA, A. et al. Evidence-based clinical practice guideline for the pharmacologic management of acute dental pain in adolescents, adults and older adults. *Journal of the American Dental Association*, v. 154, n. 12, p. 1066–1084, 2023. DOI: 10.1016/j.adaj.2023.06.014.

11. SLAYTON, R. L. et al. Evidence-based clinical practice guideline on nonrestorative treatments for carious lesions: a report from the American Dental Association. *Journal of the American Dental Association*, v. 149, n. 10, p. 837–849.e19, 2018. DOI: 10.1016/j.adaj.2018.07.002.

12. DING, H. et al. Artificial intelligence in dentistry — a review. *Frontiers in Dental Medicine*, v. 4, p. 1–20, 2023.

13. KÜNZLE, P.; PARIS, S. Performance of large language artificial intelligence models on solving restorative dentistry and endodontics student assessments. *Clinical Oral Investigations*, v. 28, artigo 575, 2024. DOI: 10.1007/s00784-024-05968-w.

14. BULDUR, M.; SEZER, B. Evaluating the accuracy of ChatGPT-4 responses to United States Food and Drug Administration frequently asked questions about dental amalgam. *BMC Oral Health*, v. 24, n. 605, 2024.

- 15.SUÁREZ, A. et al. Unveiling the ChatGPT phenomenon: evaluating the consistency and accuracy of endodontic question answers. *International Endodontic Journal*, v. 57, p. 108–113, 2024.

- 16.MONDILLO, G. et al. Basal knowledge in the field of pediatric nephrology and its enhancement following specific training of ChatGPT-4 “omni” and Gemini 1.5 Flash. *Pediatric Nephrology*, v. 40, p. 151–157, 2025.

- 17.LIM, Z. W. et al. Benchmarking large language models’ performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. *eBioMedicine*, v. 95, p. 104770, 2023.

- 18.NATURE EDITORIAL BOARD. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, v. 613, p. 620–621, 2023.

- 19.VISHWANATHAIAH, S.; FAGHEEH, H. N.; KHANAGAR, S. B.; MAGANUR, P. C. Artificial intelligence: its uses and application in pediatric dentistry – a review. *Biomedicines*, v. 11, n. 788, p. 1–19, 2023.

- 20.SCHWENDICKE, F.; SPLIETH, C.; BRESCHI, L.; BANERJEE, A.; FONTANA, M.; PARIS, S.; BURROW, M. F.; CROMBIE, F.; FOSTER PAGE, L.; GATÓN-HERNÁNDEZ, P.; GIACAMAN, R.; GUGNANI, N.; HICKEL, R.; JORDAN, R. A.; LEAL, S.; LO, E.; TASSERY, H.; THOMSON, W. M.; MANTON, D. J. When to intervene in the caries process? An expert Delphi consensus statement. *Clinical Oral Investigations*, 2019.

- 21.TEOH, LEANNE; MOSES, GERALDINE M. Are teething gels safe or even necessary for our children? A review of the safety, efficacy and use of topical lidocaine teething gels. *Journal of Paediatrics and Child Health*, v. 56, n. 3, p. 1–4, 2020.

22. MOHN, A.; MAGGITT, A.; TRIPODI, R.; CHIARELLI, F.; COSI, A.; TRIPODI, D. Dental pain in children: pharmacological management. *European Journal of Paediatric Dentistry*, v. 26, n. 2, p. 120–124, 2025. DOI: 10.23804/ejpd.2025.2190.

