



Review Article

국내외 구강보건 교육·상담용 AI 챗봇 연구 동향: 내러티브 리뷰

황수연^{ID}

대전과학기술대학교 치위생과

Research trends in artificial intelligence chatbots for oral health education and consultation: a narrative review

Su-Yeon Hwang^{ID}

Department of Dental Hygiene, Daejeon Institute of Science and Technology

Correspondence author: Su-Yeon Hwang, Department of Dental Hygiene, Daejeon Institute of Science and Technology, 100, Hyecheon-ro, Seo-gu, Daejeon, 35406, Korea. Tel: +82-42-580-6449, E-mail: hsyen@naver.com

Abstract

Objectives: This narrative review examined research trends in artificial intelligence (AI) chatbots for oral health education and counseling, with a particular focus on differences between Korean and international studies in their progress toward clinical application. **Methods:** International studies were searched in PubMed and Web of Science, and Korean studies in RISS, KCI, and DBpia, with the search limited to publications from 2019 onward. Of the 63 eligible international studies, 12 were purposively sampled to balance dental subspecialty and country of origin, with all studies reporting user or clinical outcomes retained; both Korean studies meeting the criteria were included. The 14 studies were categorized into three translational stages: development, performance evaluation, and effectiveness validation. **Results:** International studies spanned all three stages and included randomized controlled trials that assessed clinical and behavioral outcomes in patients. In contrast, the Korean studies remained at the development stage, with no assessment of response quality and no effectiveness validation involving actual users. However, many international studies were also preliminary or small in scale, indicating that widespread clinical adoption has not yet been achieved. **Conclusions:** A pronounced gap exists between Korean and international research in the clinical translation of AI chatbots for oral health education and counseling. To bridge this gap, Korean research should move beyond development toward user-centered evaluation and effectiveness validation, supported by the construction of Korean-language datasets, standardized assessment metrics, and multidisciplinary collaboration involving dental hygienists.

Keywords: Generative artificial intelligence, Health education dental, Patient education as topic

주요어: 생성형 인공지능, 구강보건교육, 환자교육

서론

구강질환은 치아우식증, 치주질환, 구강암 등을 포함하며 전 세계적으로 유병률이 높은 주요 공중보건 문제로, 개인의 건강과 삶의 질뿐만 아니라 보건의료 체계와 사회경제 전반에 상당한 부담을 초래한다[1]. 세계적으로 약 35억 명이 구강질환의 영향을 받는 것으로 추정되나, 예방적 조치와 구강보건 교육은 대부분의 구강질환에서 발생 위험을 낮추고 건강행동을 개선할 수 있다[2]. 따라서 구강질환의 예방과 관리에서 지속적이고 효과적인 구강보건 교육의 중요성이 더욱 강조되고 있다.

특히 반복적인 구강보건 교육·상담일수록 대면 방식만으로는 충분히 제공되기 어려워 이를 보완할 디지털 기반 보조도구의 필요성이 증가하고 있다. 구강보건 교육·상담은 대중이 올바른 구강위생 습관을 실천하도록 돕는 구강질환 예방의 핵심 요소이며, 치과위생사는 이를 담당하여 환자 교육, 건강상담, 동기부여 및 행동 변화 유도를 담당하는 핵심 전문가이다[3]. 그러나 대면 중심의 전통적 교육·상담은 현장 인력과 시간의 제약, 상시 피드백의 부재, 지역 간 의료 접근성의 격차 등으로 인해 모든 대상자에게 지속적이고 개별화된 서비스를 제공하기 어렵다는 한계가 있다[4].

한편 인공지능(artificial intelligence, AI)과 자연어처리(natural language processing, NLP) 기술의 발전은 의료 분야에 새로운 서비스 방식을 제공하고 있다[5]. 그 중에서도 대규모 언어모델(large language model, LLM)을 기반으로 한 AI 챗봇은 인간과 유사한 방식으로 자연어를 이해하고 생성할 수 있어, 시간과 장소의 제약 없이 개인 맞춤형 정보와 상담을 제공하는 도구로 주목받고 있다[6,7]. 기존의 모바일 앱이나 웹 기반 구강보건 교육은 사전에 제작된 콘텐츠를 일방적으로 전달하는 방식이며, 이용자의 개별 질문에 대응하거나 개인의 이해 수준에 따라 실시간으로 유연하게 응답하지 못하는 한계가 있다[8]. 반면 LLM 기반 AI 챗봇은 자연어로 표현된 질문을 이해하고 양방향 소통이 가능하며, 개인의 상황에 맞춘 실시간 맞춤형 상담을 제공한다는 점에서 디지털 매체와 차별성을 갖는다. 구강보건 영역에서도 AI 챗봇은 구강건강 지식의 전달, 개인화된 교육 콘텐츠 제공, 상담 접근성 향상 등 교육·상담의 새로운 가능성을 제시하고 있으며[9], 이러한 접근은 치과위생사의 교육·상담 업무를 보조하고 그 효율성과 형평성을 높이는 방향으로 이바지할 것으로 기대된다. 그러나 사실과 다른 내용을 생성하는 환각(hallucination) 현상, 정보의 정확성, 윤리적 문제 등은 여전히 해결해야 할 과제로 제기되고 있다[10].

이에 따라 국내외에서 구강보건 교육·상담을 위한 AI 챗봇 연구가 점차 이루어지고 있으나[11-18], 연구의 초점과 발전 단계는 국가별로 다른 양상을 보인다. 국외에서는 챗봇의 응답 정확도 평가[12,13]에서 나아가 무작위 대조시험 등 실제 이용자를 대상으로 한 효과 검증 연구[19,20]가 축적되고 있지만, 국내 연구는 챗봇을 개발하고 구현하는 단계에 주로 머물러 있어[21,22] 개발된 챗봇을 실제 임상에 적용하고 그 효과를 검증하는 단계로는 나아가지 못하고 있다. 한편 국내외 연구의 개발 단계, 연구 설계, 적용 대상 및 효과 평가 결과를 비교·분석한 문헌고찰은 매우 제한적으로 수행되어 왔다. 특히 구강보건 교육·상담 영역에서 임상 중개(clinical translation)의 발전 단계에 따라 국내외 현황과 중개 격차(translational gap)를 다룬 연구는 거의 이루어지지 않았다.

구강보건 교육·상담용 AI 챗봇 연구는 시스템 개발과 응답 정확도 평가 등 연구 설계와 결과 지표의 이질성이 높다. 따라서 본 연구는 체계적 문헌고찰보다 연구의 전반적인 흐름과 특성을 포괄적으로 검토하는데 적합한 내러티브 리뷰(narrative review)를 수행하였다. 이를 통해 연구의 주제, 대상, 연구 설계, 챗봇 유형, 평가 방법 및 효과 검증 수준을 중심으로 분석하고, 향후 국내 구강보건 교육·상담 분야에서 AI 챗봇의 개발, 임상 적용 및 효과 평가 연구의 방향성을 제시하기 위한 기초자료를 제공하고자 한다.

연구방법

1. 연구설계

본 연구는 구강보건 교육·상담용 AI 챗봇에 관한 국내외 연구 동향을 종합적으로 고찰하기 위한 내러티브 리뷰(narrative review)이다. PRISMA(Preferred Reporting Items for Systematic reviews and Meta-Analyses)지침 등 표준화된 가이드라인에 따라 문헌을 수집 평가하는 체계적 문헌 고찰과 다르게, 현재 구강보건 교육·상담용 AI 챗봇 연구는 개발 연구, 응답 정확도 평가 연구, 임상 적용 연구 등 다양한 형태로 수행되고 있어 연구설계와 결과 지표의 이질성이 높다. 따라서 본 연구는 효과 크기를 통합하는 체계적 문헌 고찰보다 연구의 발전 양상과 중개 과정을 포괄적으로 파악하기 위해 내러티브 리뷰 방법을 선택했다.

2. 문헌 검색 방법

문헌 검색은 2026년 4월 1일부터 4월 26일까지 수행하였으며, 최종 검색일은 2026년 4월 26일이다. 국외 문헌은 PubMed, Web of Science에서 직접 검색하였으며, 검색식은 (chatbot OR ChatGPT OR large language model OR artificial intelligence) AND (dental OR dentistry OR "oral health") AND (education OR consultation OR counseling)을 기본으로 설정하였다. 국내 문헌은 학술연구정보서비스(RISS), 한국학술지인용색인(KCI), DBpia를 활용하였으며, 검색식은 (ChatGPT OR 생성형 AI OR 대형 언어 모델 OR LLM OR 챗봇) AND (구강보건교육 OR 치위생 교육 OR 환자 상담 OR 치과 상담)을 적용하였다. 생성형 AI와 LLM 기반 챗봇 연구가 본격적으로 증가하기 시작한 시기를 고려하여 2019년 이후 문헌을 포함하였다.

3. 문헌 선정 및 제외 기준

선정 기준은 (1) 환자, 보호자 또는 일반 대중을 대상으로 한 구강보건 교육 또는 상담을 목적으로 AI 챗봇이나 LLM을 개발, 평가, 또는 적용한 연구 (2) 원저 논문 (3) 영어 또는 한국어로 작성되어 학술지에 게재된 문헌으로 하였다. 제외 기준은 (1) 구강 영상 진단 등 교육·상담과 직접 관련이 없는 AI 응용 연구 (2) 학술대회 초록, 사설, 단순 종설 (3) 원문 확보가 불가능한 문헌으로 하였다. 문헌의 검색과 선정은 연구자 1인이 수행하였으며, 제목과 초록을 검토한 후 전문을 확인하여 선정기준의 충족 여부를 판단하였다. 단독 연구로 수행되어 연구자 간 독립검토와의 견 불일치의 조정 절차는 적용할 수 없었으므로, 선정 및 제외 기준을 모든 문헌에 동일하게 적용하여 판단의 일관성을 유지하였다.

국외 문헌은 동일 기준 적용 시 응답 성능·정확도 평가 연구가 다수를 차지하여, 연구 유형 및 영역의 편중을 방지하고자 목적적 표집 (purposive sampling)을 실시하였다. 이에 치과 세부 진료 영역과 수행 국가별 분포를 고르게 고려하여 문헌을 배분하였으며, 편수가 제한적인 이용자·임상 효과 검증 연구는 전수 포함하여 총 12편을 선별하였다. 국내 문헌은 선정 기준을 충족한 원저 논문 2편을 별도 표집 없이 모두 포함하여, 최종 분석 대상을 총 14편으로 확정하였다. 문헌 선정 과정 및 단계별 문헌 수는 <Fig. 1>과 같다.

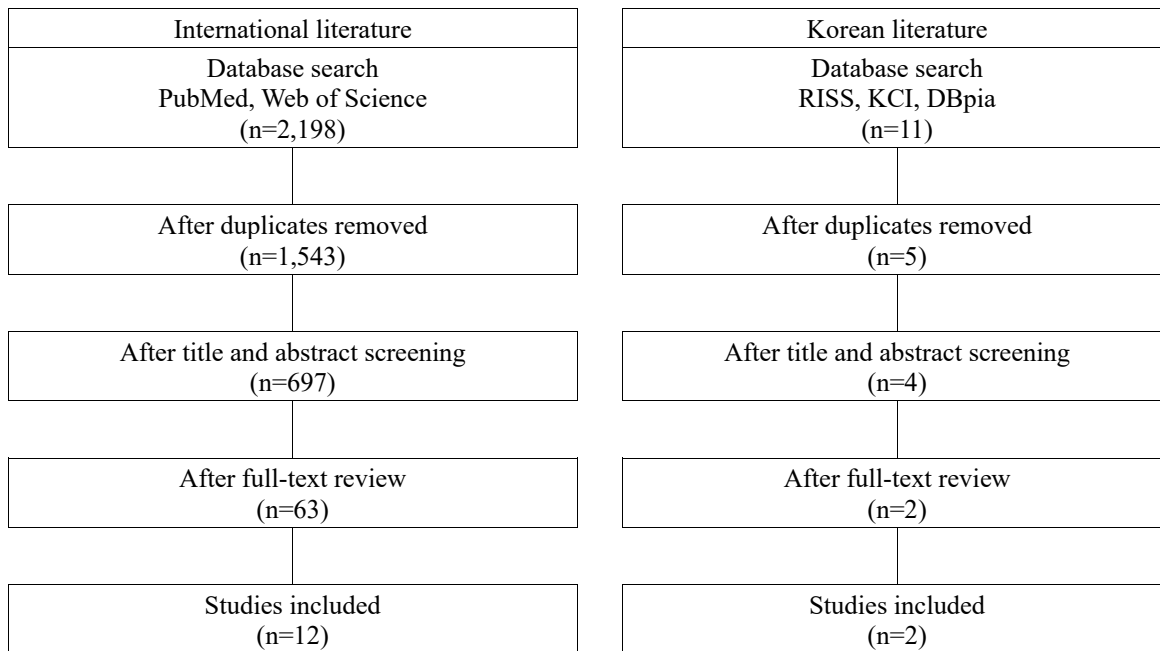


Fig. 1. Flow diagram of the literature search and selection process.

4. 문헌 분석 방법

선정된 문헌은 발표 연도, 국가, 사용된 AI 모델 및 챗봇 유형, 적용 대상, 연구 방법, 주요 결과를 기준으로 정리하였다. 특히 각 연구를 임상 중개(clinical translation)의 관점에서 (1) 개발·구축 단계 (2) 성능·정확도 평가 단계 (3) 이용자·임상 효과 검증 단계의 세 단계로 분류하여, 국내외 연구가 각 단계에 분포하는 양상을 비교·분석하였다. 이러한 구분은 의료 인공지능의 평가가 알고리즘 개발과 모의 환경에서의 성능 검증을 거쳐 실제 이용자와 임상 환경에서의 효과 검증으로 단계적으로 이행해야 한다는 평가 단계 모형에 근거하였다[23]. 본 연구에서 중개 격차(translational gap)는 AI 시스템의 개발과 실제 임상 적용 사이에 존재하는 간극을 의미하며, 국내외 연구가 위 세 단계에 분포하는 양상의 차이로 조작적으로 정의하였다. 이를 바탕으로 중개 격차의 성격과 그 원인을 서술적으로 분석하였다.

연구결과

1. 국외 구강보건 교육 및 상담용 AI 챗봇 연구 동향

국외에서는 LLM 기반 챗봇을 구강보건 교육·상담에 적용한 연구가 미국, 중국, 유럽 등 여러 국가에서 이루어지고 있으며, 시스템의 개발·구축에서 응답 성능 평가를 거쳐 실제 이용자 대상의 임상 효과 검증에 이르기까지 임상 중개(clinical translation)의 단계별로 분포했다. 본 연구에서는 선정된 국외 문헌 12편을 개발·구축 1개, 성능 평가단계 7개, 효과 검증단계 4개로 구분하여 분석했다<Table 1>.

1) 도메인 특화 모델의 개발·구축

첫 번째 단계는 특정 구강보건 영역에 특화된 모델을 개발하는 연구이다. Fanelli 등[11]은 검색증강생성(retrieval-augmented generation, RAG)을 GPT-4o에 적용하여 치주학 특화 모델 'PerioGPT'를 개발하였다. 전문가 문항으로 다섯 개의 범용 챗봇과 비교한 결과 정확도 81.16%로 가장 높은 정확도를 나타냈다. 다만 해당 연구는 제한된 문항을 기반으로 수행되었으며, 실제 이용자를 대상으로 한 평가는 포함되지 않았다.

2) LLM 응답 품질의 성능·정확도 평가

두 번째 단계는 AI 챗봇이 구강보건 질문에 생성한 응답의 품질을 전문가 평가 등으로 검증하는 연구로, 국외 문헌 7편이 이에 해당하여 가장 큰 비중을 차지하였다. 연구는 미국, 중국, 튀르키예, 독일 등 여러 국가에서 수행되었으며, 주로 ChatGPT, Bard, Gemini, Claude 등의 응답에 대한 정확성, 가독성 및 신뢰성을 평가하였다. 대표 연구를 살펴보면, 미국에서는 제3대구치 발치[12] 및 근관치료[13] 빈출 질문에 대한 응답을 지침과 비교하여 정확성은 대체로 인정되나 가독성이 낮거나 신뢰성이 충분하지 않다는 한계를 보고하였다. 중국과 튀르키예 연구진은 일반 구강질환[14], 치과교정[15], 보철·임플란트[16,17] 등 세부 치과 영역의 빈출 질문에 대한 응답 성능을 비교하여, ChatGPT 계열이 상대적으로 높은 평가를 받았으나 플랫폼 간 성능 편차와 검증할 수 없는 인용 등의 문제가 있음을 확인하였다. 독일에서는 AI가 생성한 방사선 판독 보고서를 ChatGPT로 환자 친화적으로 단순화한 결과, 무작위 배정 연구에서 환자의 이해도가 유의하게 향상되었다[18].

3) 무작위 대조시험 기반의 이용자·임상 효과 검증

세 번째 단계는 실제 이용자를 대상으로 챗봇의 효과를 임상 지표나 행동 변화로 검증하는 연구로, 국외 문헌 4편이 이에 해당하였다. 효과 검증 단계의 연구는 태국, 이탈리아, 파키스탄 등에서 수행되었으며 모두 실제 이용자를 대상으로 한 무작위 대조시험 연구였다. 대상자는 아동의 보호자, 교정환자, 근관·수복치료 환자였고, 결과 지표는 칫솔질 행동, 치은 염증 등의 임상 지표와 지식 습득, 불안 등의 이용자 반응이었다.

태국에서는 보호동기이론에 기반하여 개발된 FunDee 챗봇을 활용한 연구가 2편 확인되었다. Pithpornchaiyakul 등[19]은 챗봇이 대면 훈련 없이도 보호자의 칫솔질 기술을 향상시켰음을 보고하였고, Hunsrisakhun 등[20]은 보호자·아동을 대상으로 한 무작위 대조시험에서 독립형 챗봇이 아동의 구강건강 개선에 유효한 대안이 됨을 보고하였다. 이탈리아에서 Santonocito 등[24]은 교정환자를 대상으로 한 무작위 대조시험에서 챗봇군의 치은 염증이 유의하게 개선되었으나 지식 습득에서는 대조군과 차이가 확인되지 않았음을 보고하였다. 파키스탄에서 Islam[25]은 근관·수복치료 환자를 대상으로 AI가 생성한 교육 자료가 정보의 명료성과 신뢰를 높이고 불안을 낮추며 지식 보유를 향상시켰음을 보고하였다.

Table 1. Summary of overseas studies on AI chatbots for oral health education and consultation by translational stage

Author (year)	Country	AI model / Chatbot	Target/domain	Methods and main findings
Stage 1. Development				
Fanelli et al. (2025)	Italy	PerioGPT (periodontology-specific GPT-4o + RAG)	Periodontal knowledge / development	Built a periodontology-specific model; compared with 5 general chatbots on 50 items and validated on 71 items. Achieved the highest accuracy (81.16%), showing that a domain-specific model can outperform general LLMs.
Stage 2. Performance evaluation				
Lv et al. (2024)	China	Bard, ChatGPT	General oral disease/patients	Blinded evaluation of responses to 40 patient-derived questions by experts and laypersons. ChatGPT-4 performed best, though variability existed across models.
Jacobs et al. (2024)	USA	ChatGPT	Third molar extraction / patients	Assessed accuracy against AAOMS clinical guidelines and readability via Flesch-Kincaid. Accuracy was acceptable, but readability was low and citations were absent.
Mohammad-Rahimi et al. (2024)	USA	ChatGPT, Bard, Bing	Endodontic FAQs / patients	Evaluated the validity and reliability of chatbot responses to endodontic FAQs. Confirmed potential as an information source alongside limitations for public use.
Chen et al. (2025)	China	Ernie Bot, ChatGPT, Gemini	Orthodontic consultation / patients	Evaluated responses to 50 pre-treatment questions across 6 dimensions. Mostly scored 3–4, indicating cautious use and need for expert verification.
Özcivelek and Özcan (2025)	Türkiye	DeepSeek-R1, ChatGPT, Dental GPT	Prosthodontics / patients	Evaluated responses to 35 prosthodontic FAQs for accuracy, readability, and actionability (in-silico). Some responses contained unverifiable citations.
Tuzlalı et al. (2025)	Türkiye	ChatGPT, DeepSeek, Gemini, Claude, Perplexity	Dental implants / patients	Evaluated 45 implant FAQs across 5 chatbots. ChatGPT-o1 ranked highest (relevance 4.99), with large performance differences among platforms.
Stephan et al. (2025)	Germany	ChatGPT (report simplification)	Radiology report explanation / patients	Randomized 300 patients; the simplified version of AI-generated radiology reports improved comprehension and communication.
Stage 3. Effectiveness validation				
Pithpornchaiyakul et al. (2022)	Thailand	FunDee (dedicated chatbot)	Toothbrushing education / caregivers	Validated as an alternative to in-person training during COVID-19. The 30-day FunDee improved toothbrushing skills without face-to-face training.
Hunsrisakhun et al. (2025)	Thailand	FunDee Plus (dedicated chatbot, PMT-based)	Child oral health / caregiver-child	RCT with 320 dyads. The standalone chatbot was an effective alternative to a group combining chatbot with in-person training for improving child oral health.
Santonocito et al. (2025)	Italy	AI chatbot (guideline-based)	Orthodontic patient education/ patients	RCT with 100 orthodontic patients (chatbot vs. leaflet). Gingival inflammation (MGI) improved significantly, but knowledge gain did not differ.
Islam (2025)	Pakistan	ChatGPT (educational content)	Endodontic & restorative / patients	Randomized comparison in 100 patients. The AI group showed higher clarity, usefulness, and trust, lower anxiety, and better knowledge retention.

AI: artificial intelligence; RAG: retrieval-augmented generation; AAOMS: American Association of Oral and Maxillofacial Surgeons; LLM: large language model; FAQ: frequently asked question; RCT: randomized controlled trial; PMT: protection motivation theory; MGI: modified gingival index.

Table 2. Summary of Korean studies related to AI chatbots for oral health education and consultation

Author (year)	Category	AI technology / model	Target/data	Methods and main findings
Park (2024a)	Development	GPT + RAG (LangChain)	Dental hospital board consultation data (60,218 characters)	Built a dental consultation Q&A chatbot that retrieves consultation content similar to the user's question. Simulation suggested improved access and accuracy.
Park (2024b)	Development	RAG + function calling + AI agent	Same consultation dataset	Improved handling of complex questions via function calling and parallel processing. Simulation-stage demonstration; no user/clinical validation.

AI: artificial intelligence; GPT: generative pre-trained transformer; RAG: retrieval-augmented generation.

2. 국내 구강보건 교육·상담용 AI 챗봇 연구 동향

최종 포함된 국내 문헌은 총 2편으로, 모두 동일 연구자에 의해 수행된 개발 연구였다. 두 연구는 임상 중개 단계로는 모두 개발·구축 단계에 해당하였으며, 성능·정확도 평가 단계와 임상 효과 검증 단계에 해당하는 국내 문헌은 확인되지 않았다<Table 2>.

Park[21]은 국내 치과대학병원 게시판의 상담 내용을 수집하고 치과 전문의의 감수를 거쳐 상담 데이터셋을 구축한 뒤, GPT-3.5와 LangChain을 이용하여 RAG 기반의 치과 상담용 챗봇을 개발하였다. 사용자 질문과 유사한 상담 내용을 검색하여 응답하는 방식으로, 시뮬레이션을 통해 치과 상담의 접근성과 정확성을 높일 수 있음을 확인하였다. 이어진 후속 연구[22]에서는 동일한 데이터셋에 함수 호출(function calling)과 AI 에이전트(AI agent) 기법을 결합하여 기존 RAG가 복잡한 질문에 대응하기 어려운 한계를 보완하고자 하였으며, 역시 시뮬레이션을 통해 효과를 시연하였다.

두 연구 모두 시스템의 구축과 시뮬레이션 시연에 초점이 있어, 실제 이용자를 대상으로 한 사용성 평가와 전문가에 의한 응답 품질 평가는 수행되지 않았으며 임상 효과를 검증한 결과도 제시되지 않았다.

3. 치위생 분야 생성형 AI 관련 참고 연구

본 연구의 선정 기준에는 해당하지 않으나, 구강보건 영역에서 생성형 AI를 다룬 학생 대상 연구가 확인되었다. 한 연구에서는 대학생 198명을 대상으로 ChatGPT 이용경험과 인식을 조사하여 치위생 전공생의 이용률이 53.5%이고 주된 목적이 학습·과제임을 보고하였다[26]. 또 다른 연구에서는 치위생과 ‘치과 상담 및 진료 영어’ 교과목에서 학생이 가상 환자 역할의 AI와 상호작용하는 ChatGPT 활용 사례를 보고하여 학생의 사용 능숙도 향상과 두려움 감소를 확인하였다[27]. 이들 연구는 대상이 환자가 아닌 학생이므로 분석 대상에서 제외하고 참고 연구로 구분하였다.

총괄 및 고안

본 연구는 국내외 구강보건 교육·상담용 AI 챗봇 연구의 동향을 임상 중개(clinical translation)의 관점에서 고찰하였다. 국외 문헌 총 12편 중 1편은 도메인 특화 모델 개발, 7편은 LLM 응답 성능 평가, 4편은 실제 환자 및 보호자 대상의 임상 효과 검증 연구로 전 단계에 걸쳐 분포한 반면, 국내 문헌은 최종 선정된 2편 모두 시스템 개발·구축 단계에만 머물러 있어 국가 간 중개 격차가 정량적으로 확인되었다. 특히 국외의 경우 태국, 이탈리아, 파키스탄 등에서 수행된 무작위 대조시험(RCT)을 통해 챗봇 적용 후 칫솔질 수행 기술 향상, 치은 염증 지수(MGI)의 유의한 개선, 환자의 교육 만족도 향상 및 시술 불안감 감소 등 실제적인 임상·행동 지표의 효과를 입증하였다. 반면 국내 연구 2편은 치과 대학병원 상담 데이터 기반의 RAG 및 AI 에이전트 구축과 단순 시뮬레이션 시연에 한정되어 있어, 응답 품질에 대한 정밀 평가나 실제 이용자 대상의 효과 검증은 수행되지 않았다.

다만 이러한 격차가 곧 국외에서 AI 챗봇이 구강보건 교육·상담에 정착되었음을 의미하지는 않는다. 국외 연구 또한 대체로 소규모·예비적 단계에 머물러 있고, 임상 지표는 개선되었으나 지식 습득에서는 대조군과 차이를 보이지 못한 경우가 있으며[24], 응답의 정확성·가독성·신뢰성에 대한 우려도 공통으로 지적된다[11,14]. 이는 챗봇이 행동 변화나 접근성 향상에는 이바지할 수 있으나, 지식 전달 자체에서는 기존 교육 자료와 유사한 수준일 가능성을 시사한다.

국내 연구가 개발 단계에 머무는 배경에는 몇 가지 요인이 작용한다. 첫째, 한국어에 특화된 치의학·구강보건 언어 자원의 부족이다. 의료 분야 전반에서도 한국어 언어 자원의 부족이 지속적으로 제기되어 왔으며[28], 국외에서는 치주학 특화 모델(PerioGPT)처럼 도메인 특화 데이터로 성능을 높인 사례가 등장하고 있으나[11], 국내에는 환자 교육·상담에 활용할 표준화된 한국어 데이터셋이 구축되었다는 보고가 확인되지 않았다. 실제로 본 연구에서 분석한 국내 문헌은 모두 상담 게시판 자료를 자체적으로 수집하여 데이터셋을 구축하였다[21,22]. 둘째, 구강보건 데이터의 표준화와 접근성 문제로, 데이터의 형식이 다양하고 개인정보 보호의 제약이 있어 학습·검증용 자료의 확보가 쉽지 않다. 셋째, 임상 검증에 필요한 학제간 협력과 제도적 기반의 부족이다. 이용자 대상 효과 검증은 임상 전문가와 AI 개발자, 연구윤리·규제 체계의 협력을 필요로 하나 국내에서는 이러한 구조가 아직 충분히 마련되지 않았으며, AI가 생성한 정보의 정확성과 책임 소재 등 윤리적 쟁점 또한 임상 적용의 문턱을 높인다. 넷째, 교육 콘텐츠의 전문적 검증 체계가 마련되어 있지 않다. 구강보건 교육·상담용 AI 챗봇의 응답은 기술적 정확성뿐 아니라 교육 내용으로서의 타당성을 갖추어야 하나, 국내 연구는 상담 게시판의 데이터를 활용하는데 그쳐[21,22] 구강보건교육의 표준안에 근거한 콘텐츠의 구성과 검증은 이루어지지 않았다. 구강보건교육을 수행해 온 치과위생사가 데이터셋의 구성과 응답 내용의 타당도 검증에 참여함으로써, 전문적으로 검증된 교육 콘텐츠를 구축할 필요가 있다.

한편 효과 검증 단계에 도달한 연구들 사이에서도 중재의 설계 방식에 따라 결과가 달랐다. 보호동기이론(Protection Motivation Theory)에 기반하여 설계된 AI 챗봇은 아동의 구강건강 개선이라는 임상 지표의 변화를 확인하였다[20]. 그러나 정보 제공에 초점을 둔 AI 챗봇을 적용한 연구에서는 지식 습득에서 대조군과 차이가 확인되지 않았다[24]. 이는 정보 제공 기능만으로 지속적인 행동 변화를 유도하기 어려우며, 교육학적·행동과학적 근거가 함께 적용되어야 함을 의미한다. 따라서 국내에서도 기술 개발 중심의 접근을 넘어 행동변화 이론을 반영한 구강보건 상담용 AI 챗봇의 개발 연구가 필요하다.

본 연구는 구강보건 교육·상담이라는 특정 영역에 초점을 맞추어 국내외 연구를 직접 비교하고, 이를 임상 중개의 세 단계로 구조화하여 발전 단계의 차이를 드러냈다는 점에서 의의가 있다. 다만 내러티브 리뷰의 특성상 체계적 문헌고찰과 달리 문헌의 질 평가(quality assessment)를 수행하지 못하여 개별 연구의 방법론적 수준을 반영하지 못하였고, 이질적인 연구 유형을 포괄하였으므로 정량적 메타분석을 통한 효과의 통합도 이루어지지 않았다. 또한 문헌의 검색과 선정을 단독 연구자가 수행하여 연구자 간 독립 검토와 의견 불일치의 조정 절차를 거치지 못하였다. 이는 문헌의 선정과 해석에 연구자의 판단이 개입될 여지를 남기며 선택편향(selection bias)의 가능성을 높이는 요인이 된다. 국외 문헌을 목적적 표집으로 선별하는 과정에서 일부 문헌이 포함되지 않았을 수 있다.

본 연구의 결과는 국내 구강보건 교육·상담 분야에 다음과 같은 함의를 제공한다. AI 챗봇은 치과위생사의 교육·상담 업무를 대체하는 것이 아니라 보조하는 도구로서, 시간과 장소의 제약을 넘어 개인 맞춤형 교육과 반복 학습을 지원하고 접근성과 형평성을 높이는데 이바지할 수 있다[20,25]. 효과 검증 단계에 도달한 연구들이 AI 챗봇을 대면 교육의 대체재가 아니라 이를 보완하는 형태로 설계하였다는 점은 AI 챗봇이 전문가가 제공하는 교육을 보완하는 형태에서 가장 효과적으로 활용될 수 있음을 시사한다.

AI 챗봇의 개발과 검증 과정에서 치과위생사는 AI 챗봇이 다룰 내용의 범위와 대화 시나리오를 설계하고 데이터셋의 내용 타당도를 검증하며, 대상자의 이해 수준에 적합한지를 판단하는 평가 준거를 마련하는 역할을 수행할 수 있다. 또한 응답의 정확성·가독성·신뢰성에 한계가 있으므로[12,13] AI가 생성한 정보를 검토하고 보완하는 역할도 요구된다. 나아가 국내에서도 실제 이용자를 대상으로 한 응답 품질 평가와 임상 효과 검증 연구가 요구되며, 이를 위해 한국어 구강보건 데이터셋의 구축, 표준화된 평가 지표의 마련, 치의학·치위생 임상과 AI 개발의 학제간 협력과 함께 치과위생(학)과 교육과정과 보수교육에서의 AI 활용 역량 함양이 병행될 필요가 있다.

결론

본 연구는 국내외 구강보건 교육·상담용 AI 챗봇 연구의 동향을 임상 중개의 관점에서 고찰한 내러티브 리뷰로, 국외 문헌 12편과 국내 문헌 2편을 개발·구축, 성능·정확도 평가, 이용자·임상 효과 검증의 세 단계로 분류하여 분석하였다. 그 결과 국외 연구는 개발 단계에서 실제 이용자를 대상으로 한 효과 검증 단계까지 발전하였다. 하지만 국내 연구는 시스템 개발 단계에 집중되어 있어 임상 중개 수준에서 차이가 확인되었다. 다만 국외 연구 또한 대체로 소규모·예비적 단계로 실제 임상과 지역사회로의 대중적 적용에는 이르지 못하여, 국외에서도 AI 챗봇의 효과와 안전성을 일반화하기 위한 추가 검증이 요구된다.

따라서 국내에서도 개발 단계를 넘어 실제 이용자를 대상으로 한 응답 품질 평가와 임상 효과 검증 연구가 요구되며, 이를 위해 한국어 구강보건 데이터셋의 구축, 표준화된 평가 지표의 마련, 치의학·치위생 임상과 AI 개발의 학제간 협력이 필요하다. 특히 신뢰성 있는 한국어 구강보건 데이터셋의 구축은 향후 국내 AI 챗봇 연구의 기반이 되는 핵심 과제로 판단된다. 구강보건 교육과 상담의 주요 수행자인 치과위생사는 AI 챗봇을 활용한 디지털 구강보건교육의 설계, 운영 및 평가 과정에서 중요한 역할을 수행할 수 있다.

본 연구 결과는 향후 AI 기반 구강보건교육 및 상담 서비스의 개발 방향을 제시하고, 치과위생사의 디지털 교육역량 강화와 근거기반 챗봇 활용 전략 수립을 위한 기초자료로 활용될 수 있을 것이다. 다만 본 연구는 내러티브 리뷰의 특성상 문헌 선정과 해석 과정에 연구자의 판단이 개입될 수 있으므로 결과 해석에 주의가 필요하다.

Notes

Author contributions

The author fully participated in the work performed and documented truthfully.

Conflicts of interest

The author declared no conflicts of interest.

Funding

None.

Ethical statement

This study is a literature review that did not involve human participants, human-derived materials, or animals. Therefore, approval by an Institutional Review Board was not required.

Data availability

The datasets generated during the current study are available from the corresponding author upon request.

Acknowledgments

None.

Use of artificial intelligence (AI) tools and software

No AI tools or AI-assisted technologies were used in the preparation of this manuscript.

References

1. Peres MA, Macpherson LMD, Weyant RJ, Daly B, Venturelli R, Mathur MR, et al. Oral diseases: a global public health challenge. *Lancet* 2019;394(10194):249–60. [https://doi.org/10.1016/S0140-6736\(19\)31146-8](https://doi.org/10.1016/S0140-6736(19)31146-8)
2. World Health Organization. Global oral health status report: towards universal health coverage for oral health by 2030. Geneva: World Health Organization. 2022.
3. Kwun HS, Jo GS. A study on the role of dental hygienist for revitalization of dental health class in community health center. *J Korean Acad Dent Hyg Educ* 2006;6(4):263–82.
4. Han KS, Lee SJ, Shin JH. Non-contact nursing education in communication and group counseling: current status and future directions. *Stress* 2023;31(4):149–57. <https://doi.org/10.17547/kjsr.2023.31.4.149>
5. Yoon SO. Issues of public service using artificial intelligence: focused on chatbot service. *Korean Public Manag Rev* 2018;32(2):83–104. <https://doi.org/10.24210/kapm.2018.32.2.004>
6. Jung J. Client's psychological counseling experience using ChatGPT: focusing on phenomenological research method. *Discourse Policy Soc Sci* 2025;18(2):229–62. <https://doi.org/10.22417/DPSS.2025.10.18.2.229>

7. Yang EY. Necessity of regulation on the development and use of generative AI: focusing on large language models conversational A.I. services (LLMs AI). *Sungkyunkwan Law Rev* 2023;35(2):293–325. <https://doi.org/10.17008/skklr.2023.35.2.009>
8. Väyrynen E, Hakola S, Keski-Salmi A, Jämsä H, Vainionpää R, Karki S. The use of patient-oriented mobile phone apps in oral health: scoping review. *JMIR Mhealth Uhealth* 2023;11:e46143. <https://doi.org/10.2196/46143>
9. Li H, Chen S, Chang B, Wang X, He Y, Xu B, et al. Application of artificial intelligence in oral health management: challenges and opportunities. *Front Med (Lausanne)* 2026;13:1700529. <https://doi.org/10.3389/fmed.2026.1700529>
10. Berk H. Beware of artificial intelligence hallucinations or should we call confabulation? *Acta Orthop Traumatol Turc* 2024;58(1):1–3. <https://doi.org/10.5152/j.aott.2024.130224>
11. Fanelli F, Saleh M, Santamaria P, Zhurakivska K, Nibali L, Troiano G. Development and comparative evaluation of a reinstructed GPT-4o model specialized in periodontology. *J Clin Periodontol* 2025;52(5):707–16. <https://doi.org/10.1111/jcpe.14101>
12. Jacobs T, Shaari A, Gazonas CB, Ziccardi VB. Is ChatGPT an accurate and readable patient aid for third molar extractions? *J Oral Maxillofac Surg* 2024;82(10):1239–45. <https://doi.org/10.1016/j.joms.2024.06.177>
13. Mohammad-Rahimi H, Ourang SA, Pourhoseingholi MA, Dianat O, Dummer PMH, Nosrat A. Validity and reliability of artificial intelligence chatbots as public sources of information on endodontics. *Int Endod J* 2024;57(3):305–14. <https://doi.org/10.1111/iej.14014>
14. Lv X, Zhang X, Li Y, Ding X, Lai H, Shi J. Leveraging large language models for improved patient access and self-management: assessor-blinded comparison between expert- and AI-generated content. *J Med Internet Res* 2024;26:e55847. <https://doi.org/10.2196/55847>
15. Chen J, Ge X, Yuan C, Chen Y, Li X, Zhang X, et al. Comparing orthodontic pre-treatment information provided by large language models. *BMC Oral Health* 2025;25(1):838. <https://doi.org/10.1186/s12903-025-06246-1>
16. Özcivelek T, Özcan B. Comparative evaluation of responses from DeepSeek-R1, ChatGPT-o1, ChatGPT-4, and dental GPT chatbots to patient inquiries about dental and maxillofacial prostheses. *BMC Oral Health* 2025;25(1):871. <https://doi.org/10.1186/s12903-025-06267-w>
17. Tuzlali M, Baki N, Aral K, Aral CA, Bahçe E. Evaluating the performance of AI chatbots in responding to dental implant FAQs: a comparative study. *BMC Oral Health* 2025;25(1):1548. <https://doi.org/10.1186/s12903-025-06863-w>
18. Stephan D, Bertsch AS, Schumacher S, Puladi B, Burwinkel M, Al-Nawas B, et al. Improving patient communication by simplifying AI-generated dental radiology reports with ChatGPT: comparative study. *J Med Internet Res* 2025;27:e73337. <https://doi.org/10.2196/73337>
19. Pithpornchaiyakul S, Naorungroj S, Pupong K, Hunsrisakhun J. Using a chatbot as an alternative approach for in-person toothbrushing training during the COVID-19 pandemic: comparative study. *J Med Internet Res* 2022;24(10):e39218. <https://doi.org/10.2196/39218>
20. Hunsrisakhun J, Naorungroj S, Tangkuptanon W, Wattanasit P, Pupong K, Pithpornchaiyakul S. Impact of oral health chatbot with and without toothbrushing training on childhood caries. *Int Dent J* 2025;75(2):1348–59. <https://doi.org/10.1016/j.identj.2024.09.028>
21. Park J. Development of dental consultation chatbot using retrieval augmented LLM. *J Inst Internet Broadcast Commun* 2024;24(2):87–92. <https://doi.org/10.7236/JIIBC.2024.24.2.87>
22. Park J. Development of chatbot for dental consultation using AI agent. *J Inst Internet Broadcast Commun* 2024;24(6):217–22. <https://doi.org/10.7236/JIIBC.2024.24.6.217>
23. Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: phases of clinical research. *JAMIA Open* 2020;3(3):326–31. <https://doi.org/10.1093/jamiaopen/ooaa033>
24. Santonocito S, Cicciù M, Ronsivalle V. Evaluation of the impact of AI-based chatbot on orthodontic patient education: a preliminary randomised controlled trial. *Clin Oral Investig* 2025;29(5):278. <https://doi.org/10.1007/s00784-025-06356-8>
25. Islam S. Evaluating the impact of AI-generated educational content on patient understanding and anxiety in endodontics and restorative dentistry: a comparative study. *BMC Oral Health* 2025;25:689. <https://doi.org/10.1186/s12903-025-06069-0>
26. Hwang YS, Joo SJ, Kim MH. ChatGPT usage experience and perception among dental hygiene students. *J Korean Acad Dent Adm* 2024;12(1):124–32. <https://doi.org/10.22671/JKADA.2024.12.1.124>
27. Choi YK, Kim MS. A case study on the use of ChatGPT in dental English education: exploring AI-based teaching and learning methods. *J Korean Dent Hyg Sci* 2025;8(1):49–58. <https://doi.org/10.22753/JKDHS/2025.8.1.49>
28. Byun SJ, Hong JS, Park SM, Jang DJ, Seo J, Kim MS, et al. Korean bio-medical corpus (KBMC) for medical named entity recognition. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*; 2024 May 20–25; Torino, Italy. ELRA and ICCL; 2024: 9941–7.