

大语言模型在口腔医学的应用

冯钺煜, 朱海华*

浙江大学医学院附属口腔医院, 浙江大学口腔医学院, 浙江省口腔疾病临床医学研究中心, 浙江省口腔生物医学研究重点实验室, 浙江 杭州

收稿日期: 2026年4月12日; 录用日期: 2026年5月6日; 发布日期: 2026年5月12日

摘要

目的: 系统梳理大语言模型(Large language models, LLMs)在口腔医学领域的应用现状、技术进展与未来发展方向, 为口腔医学的智能化转型提供参考。方法: 通过文献综述与分析, 从临床诊断与治疗规划、患者教育与沟通、口腔医学教育三个核心应用领域, 系统归纳LLMs的应用效果与性能表现, 并深入探讨其在口腔医学应用中面临的挑战与未来发展方向。结果: LLMs在口腔医学多个领域展现出巨大潜力。在临床诊断方面, 通过微调可显著提升诊断准确性(如多生牙检测准确率从63%提升至91%), 在病理诊断中与专家一致性最高可达68.6%; 在治疗规划中, 模型在遵循临床指南方面准确率可达80%, 但整体性能仍不稳定且存在异质性。在患者教育与沟通方面, LLMs回答问题的综合准确率为81.87%, 并能生成科学可靠的科普信息与知情同意材料。在口腔医学教育中, LLMs作为辅助工具, 在多项目牙科考试中的表现已达到甚至超过考生平均水平。然而, 当前应用面临准确性波动、“幻觉”现象、数据隐私风险、模型“黑箱”问题以及缺乏标准化评估框架等主要挑战。结论: LLMs在口腔医学中具有广阔的应用前景, 其定位应是辅助而非替代临床决策。未来, 通过发展多模态模型、建立统一评估基准、增强模型可解释性以及开展前瞻性临床验证, LLMs有望成为可靠的医生助手, 助力口腔医学服务体系向更高效、高质量的方向发展。

关键词

大语言模型, 口腔医学, 人工智能

Application of Large Language Models in Stomatology

Chengyu Feng, Haihua Zhu*

Stomatology Hospital, School of Stomatology, Zhejiang University School of Medicine, Zhejiang Provincial Clinical Research Center for Oral Diseases, Zhejiang Key Laboratory of Oral Biomedical Research, Hangzhou Zhejiang

Received: April 12, 2026; accepted: May 6, 2026; published: May 12, 2026

*通讯作者。

Abstract

Objective: This paper aims to systematically review the current state, technological advancements, and future directions of large language models (LLMs) in the field of stomatology, providing insights for the intelligent transformation of oral healthcare. **Methods:** A comprehensive literature review was conducted to synthesize applications and performance of LLMs across three core domains: clinical diagnosis and treatment planning, patient education and communication, and dental education. The analysis also focused on identifying the primary challenges and outlining future development pathways for integrating LLMs into stomatology. **Results:** LLMs have demonstrated significant potential across multiple facets of stomatology. In clinical diagnosis, domain-specific fine-tuning substantially improved performance, for instance, increasing the accuracy of supernumerary tooth detection from 63% to 91%. In pathological diagnosis, the highest level of agreement with experts reached 68.6%. For treatment planning, LLMs showed an 80% accuracy in adhering to clinical guidelines; however, overall performance remained inconsistent and was assessed using heterogeneous methods. In patient education and communication, LLMs achieved a pooled accuracy of 81.87% in answering patient inquiries and generated scientifically reliable information and informed consent materials. Within dental education, LLMs served as powerful adjunct tools, performing at or above the average level of human examinees on various dental board examinations. Despite these promising results, several critical challenges persist, including performance variability, the propensity to generate inaccurate “hallucinations”, data privacy concerns, the inherent “black box” nature of the models, and the absence of a standardized evaluation framework. **Conclusion:** LLMs hold immense promise for stomatology, positioned as assistive tools to augment rather than replace clinical expertise. Future progress hinges on the development of multimodal models, the establishment of unified and reliable benchmarks, advancements in model interpretability, and rigorous prospective clinical validation. With these advances, LLMs are poised to evolve into reliable assistants for clinicians, ultimately facilitating a more efficient and high-quality oral healthcare delivery system.

Keywords

Large Language Models, Stomatology, Artificial Intelligence

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

大语言模型(Large language models, LLMs)是一种基于海量文本数据训练的深度学习模型,其能够理解和生成自然语言,执行复杂推理任务,并在多种专业领域展现出接近人类水平的性能[1]。自2022年OpenAI发布ChatGPT以来,LLMs在全球范围内引发了广泛关注和应用热潮,其在医疗健康领域的潜力也迅速成为研究热点[2]。

口腔医学作为临床医学的重要分支,正经历着从传统临床经验驱动模式向数据驱动、智能化模式的转型。我国持续面临口腔疾病发病率居高不下、高质量口腔诊疗资源分布不均等问题,亟需新技术赋能医疗服务体系[3]。在此背景下,LLMs凭借其强大的自然语言处理能力和知识整合能力,在口腔医学的临床诊疗、患者沟通、医学教育、科学研究和医院管理等多个环节展现出广阔应用前景[4]。

目前,LLMs在口腔医学领域的应用研究呈现快速增长态势。Ronsiville等的系统综述显示,2023年

至 2024 年间发表的口腔颌面外科相关文献已达 49 篇, 但符合严格纳入标准的高质量研究仍有限[5]。Sahoo 等人在牙体牙髓和口腔修复学的系统综述则共纳入了 11 项研究, 并证实 GPT-4o 在临床诊断中展现出 72%~99% 的准确率。这些研究表明, LLMs 正在深度融入口腔医学的各个亚专业领域, 且取得了大量研究成果[6]。

本文旨在系统梳理 LLMs 在口腔医学领域的应用现状、技术进展与未来方向。并将从 LLMs 在口腔医学临床诊断与治疗规划、患者教育与沟通, 口腔医学教育和挑战与展望四个方向进行介绍。

2. 大语言模型在口腔医学临床诊断与治疗规划中的应用

2.1. 口腔病理学与影像学诊断

LLMs 在口腔疾病诊断中的辅助作用已成为研究热点。组织病理是诊断许多口腔疾病, 尤其是口腔颌面部肿瘤的金标准。但即使是经验丰富的临床医生, 也难以得到理想的诊断准确率[7]。因此, 探索 LLMs 在口腔病理学辅助诊断中的价值具有重要意义。Tassoker 使用 GPT-4o 对 123 例具有挑战性的口腔颌面部病变病例进行诊断, 最终得到高达 78% 准确率的结果, 并且误诊案例中, GPT-4o 也可以提出许多合理的鉴别诊断[8]。为了评估 LLMs 在口腔病理学决策的性能, Alvarez-Silberberg 等人对四种主流 LLM (ChatGPT-4.0、ChatGPT o1-preview、Gemini 和 Meta AI) 进行了系统评估。研究结果显示, ChatGPT o1-preview 在病理学诊断中的表现最为突出, 68.6% 的输出与专家诊断一致, 与病理学家达到“高度一致” ($\kappa = 0.66$); 而 Gemini 表现较差 (27.5%, $\kappa = 0.24$) [9]。Mohammad-Rahimi 等人评估了不同 LLMs 在回答口腔病理学中复杂问题的表现, 并发现 Claude 获得了最高的平均分[10]。

口腔颌面放射学是 LLMs 应用的另一个重要领域。Alvarez-Silberberg 等的比较研究评估了 ChatGPT-4、DeepSeek-R1、Google Gemini 和 Microsoft Copilot 在回答口腔颌面放射学多选题中的表现。结果发现, 在文本题中, ChatGPT 准确率最高 (90.5%); 在图像题中 ChatGPT 仍保持 90.0% 的准确率, 并且在不同专业主题中, 唾液腺疾病相关问题的准确率最低[9]。Aşar 等人通过开发定制化 GPT-4V 模型用于根尖片上的多生牙检测, 通过专业文献和标注图像训练后, 使模型准确率从 63% 提升至 91%, 假阳性率从 42% 降至 16%。这一结果充分表明, 通过特定领域的微调可以显著提升 LLMs 的诊断性能[11]。Kim 等人评估了 GPT-4o 和 Gemini 2.5 Pro 在颌骨病变选择题回答中的表现, ChatGPT 在结合全景片、CT 与病理时准确率达 82%, 显著优于 Gemini [12]。

上述研究表明, 在口腔临床医学的辅助诊断方面, LLMs 已经表现出巨大的潜力, 尤其是进行微调后, 模型性能能得到较大的提升。

2.2. 治疗规划与临床决策支持

LLMs 在辅助制定治疗计划方面展现出潜力。通过整合患者病史、临床检查和影像学发现, LLMs 能够生成个性化的治疗方案建议, 并引用临床指南和最新研究提供循证推荐。然而, 目前的研究多局限于理论验证层面, 尚缺乏前瞻性临床研究证实 LLMs 辅助治疗规划的实效性[13]。

Makrygiannakis 等人比较 ChatGPT、Google Bard 和 Microsoft Bing 在正畸学循证问题中的表现。发现 Microsoft Bing Chat 得到了最高的平均分, 但是所有模型都会出现缺乏全面性、准确性或相关性的答案, 提示 LLMs 模型仍存在做出错误医疗决策风险[14]。Rewthamrongsris 等人则评估了 LLMs 在口腔操作中基于感染性心内膜炎预防建议的准确性, 并且得到了 80% 的准确性, 展现出 LLMs 在遵循临床指南方面巨大的潜力[15]。Tokgöz Kaplan 和 Cankar 比较了 ChatGPT 与 Gemini 在牙脱位处理中的循证表现, 发现 Gemini 平均得分在统计学上显著高于 ChatGPT, 这个结果也证实在牙脱位上 LLMs 具有较好的可依赖性[16]。Camargo 等人的发现则发现 LLMs 在磨牙症的表现中等, 表明这些工具目前仍不能代替专业的口腔指导[17]。

综上, 在口腔领域, 尽管 LLMs 展现出巨大潜力, 但其目前的性能仍不足以替代口腔医生的专业判断, 临床医生仍然需要谨慎看待其给予的治疗方案与临床决策。另外, 现有的评估方法存在较大异质性, 包括二元评估、Likert 量表和专家共识等多种方式, 需要建立标准化的用于 LLMs 临床性能的评估框架。

3. 大语言模型在患者教育与沟通中的应用

LLMs 在患者教育领域的应用是最早被探索的方向之一。Zhang 等的系统综述和 Meta 分析专门评估了 ChatGPT 及其他 LLMs 在回答牙科患者问题中的应用效果。研究纳入 25 项发表于 2022 年 11 月至 2024 年 6 月间的原始研究, 合并分析显示 LLMs 回答的准确率为 81.87%, 其中 69.9% 的回应被认为临床可接受。亚组分析显示, ChatGPT-3.5 的准确率显著高于 Microsoft Bing, 但与 ChatGPT-4.0 和 Google Bard 无显著差异[18]。Helvacioğlu-Yigit 等人的研究评估了 LLMs 在口腔颌面外科患者教育中的应用效果, 发现 LLMs 对患者常见问题的回答具有较好的科学性与可读性[19]。Çitir 评估 ChatGPT 在口腔癌信息提供中的可靠性, 发现其给出的答案大部分都得到了良好及以上的评价, 体现了 ChatGPT 在口腔癌信息方面具有较好的可靠性[20]。Aguiar de Sousa 评估 ChatGPT 作为下颌第三磨牙手术科学信息来源的可靠性, 发现 ChatGPT 提供的 90.60% 的回答都是安全及准确的, 使其可被视作是一个可靠且易于使用的平台[21]。

LLMs 在患者教育和健康信息提供方面具有许多优势, 比如相较于人类医生, 其可以提供 24 小时可及的健康信息; 另外, 其可以针对患者的不同问题与询问方式, 生成通俗易懂、个性化的解释内容; 对于语言多样化较高的地区, LLMs 支持多语言交流, 有助于克服语言障碍[22]。最后, LLMs 可以模拟对话式交互, 使患者通过自然语言描述症状, 系统提供初步的病情分析和就诊建议, 不仅可以提升患者参与度, 也可以使口腔医疗资源匮乏的偏远地区获得及时可取的口腔健康知识, 缩小城乡口腔健康服务差距。北京大学口腔医院的实践表明, 基于 LLMs 的智能体已广泛应用于智能分诊和预问诊生成, 有效提升了医疗服务效率[3]。

另外, 在口腔医生进行口腔种植、颌面外科手术等侵入性治疗前, 往往需要与患者沟通, 取得患者的知情同意。LLMs 可以辅助生成结构清晰、语言通俗的知情同意材料, 帮助患者理解治疗过程、风险和替代方案。此外, LLMs 能够模拟医患对话场景, 帮助年轻医生练习沟通技巧, 提升共情表达能力[4]。

然而, 现有研究提示, LLMs 在应对复杂临床情况和处理患者情感敏感问题时仍存在局限, 必须设置明确的安全边界和人工复核机制[5]。

4. 大语言模型在口腔医学教育中的应用

LLMs 正在深刻改变口腔医学教育的模式。对于口腔医学生和年轻医师, LLMs 可作为强大的学习辅助工具, 提供即时信息访问和概念解释。Nguyen 等的研究指出, GPT-4 模型在多项国家牙科委员会考试题目中的得分超过及格线, 而学生普遍使用 ChatGPT 进行辅助厘清概念和考试前复习[23]。

Claman 等人认为, LLMs 可根据学生个体需求调整教学材料: 其既可以对口腔医学生提供基础概念解释, 也可以对年轻的口腔医生推送最新研究进展; 另外, 其可以生成模拟病例用于病例导向学习; 也可以根据指南或教科书创建自测题目帮助学生巩固知识[24] [25]。

另外, 基于 LLMs 的对话可以用于模拟标准化患者, 供学生练习病史采集、医患沟通和临床推理。这种虚拟训练方式具有无风险、可重复、可定制等优势, 学生可在安全环境中反复练习, 直至掌握核心技能[26]。

LLMs 在口腔医学考试中的表现已成为评估其专业能力的重要指标。多项研究证实, GPT-4 类模型在牙科执照考试和专科住院医师考试中的得分已达到甚至高于平均考生水平[27]-[31]。这一方面验证了模型的知识储备和推理能力, 另一方面也提示考试形式和内容可能需要相应调整, 以更好评估学生的临床

综合能力而非知识记忆。

5. 大语言模型的挑战与展望

5.1. 大语言模型在口腔医学应用的挑战

LLMs 在实际应用中面临的首要挑战是准确性问题。尽管在大部分标准化测试中表现优异, LLMs 在真实临床场景中的表现仍有较大波动。模型可能生成表面上合理但实际错误的信息, 即“幻觉”现象[32]。口腔医学是典型的多模态学科, 临床决策高度依赖影像学资料、口内扫描数据、三维模型与病历文本的综合分析。“幻觉”现象在口腔医学领域可能表现为虚构患者病史, 引用不存在或错误的研究结论, 或者是推荐明显错误的治疗程序。中文 LLMs 在口腔种植学任务中的评估显示, 开放性问题幻觉率(23.9%)显著高于选择题(12.45%), 而长文本建模策略可将幻觉率降低近 30% [33]。更为关键的是, LLM 给予的“幻觉”信息可能直接导致不当的种植体位置设计、正畸方案偏差或手术规划错误, 其后果较一般文本错误更为严重[34]。

另外, 由于大多数 LLMs 由商业公司运营, 其数据处理和存储政策并不透明。因此, 将患者数据输入商业 LLMs 平台存在严重的隐私泄露风险[35]。在口腔临床场景中, 隐私风险不仅涉及文本病历, 还涵盖口内扫描数据、X 线片影像、面部三维照片等高度敏感的生物特征信息, 这些数据的泄露风险远高于纯文本数据。这要求医疗机构在使用 LLMs 时必须严格遵守数据保护法规, 必要时采用本地部署或私有化解决方案[36]。当前已有研究探索联邦学习在口腔影像隐私保护中的应用, 可在不共享原始数据的前提下实现多中心协同建模[37]。另外, 也有学者提出了具有保护隐私功能的 LLMs 智能检索系统, 这使得隐私泄露风险得到了降低[38]。

“黑箱”问题是 LLMs 临床应用的重大障碍。这意味着人们难以理解模型为何得出特定的答案, 这大大影响了医生和患者对模型决策的信任度[39]。在口腔颌面外科领域中, 理想的 LLMs 应该是“可解释的 AI”, 其应该提供可解释的输出或显示影响预测的关键因素, 从而使口腔颌面外科医生能整合其建议并做出正确的临床判断[40]。另外, LLMs 在复杂临床场景和处理患者情感敏感互动中的表现仍然受限, 需要研究者们进一步改进提示工程、进行更多临床验证和伦理监管介入[41]。

5.2. 大语言模型在口腔医学应用未来的发展方向

当前 LLMs 在纯文本任务中表现优异, 但在视觉依赖型任务中仍有较大提升空间。因此, 后续研究者们需要进一步探索多模态 LLMs 的潜能[42]。多模态 LLMs 的发展能够更自然地整合文本、影像、3D 扫描等数据, 实现更准确和全面的临床决策支持。

其次, 对于评估生成式人工智能的性能, 建立一个统一、可靠的基准至关重要。这些基准不仅可以用于进行模型性能的比较, 也对未来建立行业统一标准有着重要作用[43]。

另外, 为了减少 LLMs 内部机制的不透明性, 降低其后续应用的不透明性, 需要增加模型的可解释性[44]。目前的研究数据来源多为单中心数据, 为了提升模型的泛化能力和代表性, 后续需要在保护患者隐私的前提下, 实现多中心数据协同建模, 同时开展设计严谨的前瞻性临床试验, 以验证 LLMs 辅助的真实临床价值。

最后, 因为口腔治疗多为不可逆操作, 错误决策的后果直接且难以挽回。所以, 亟需建立适应人工智能辅助决策特点的医疗责任认定机制, 并明确 LLMs 在口腔医学实践中的法律地位。部分法律学者认为, 在责任归属的判定上, 需区分 LLMs 介入的具体方式: 若其仅提供辅助建议而由医师独立决策, 则医师对最终方案负主要责任; 若医师实质上依赖 LLMs 输出而未进行独立判断, 则需考虑其系统缺陷与医师过失的竞合[45]。

6. 结论

LLMs 技术正在深度融入口腔医学的各个领域,从临床辅助诊断、治疗规划、患者教育,到医学教育,应用场景不断拓展。

然而,LLMs 在口腔医学领域的定位应是辅助而非替代临床决策。当前模型仍面临准确性不足、“幻觉”风险、数据隐私担忧、泛化能力有限等挑战,尚不具备作为独立诊断工具的条件。临床实践中必须保持人类医生的最终决策权和监督责任。

技术的临床价值最终取决于能否与医师的专业判断形成有效协同,在提高诊疗效率和可及性的同时,确保患者安全和医疗质量。未来,随着多模态融合、领域优化、可解释性算法和前瞻性临床验证的推进,LLMs 有望从单纯的辅助工具变成可靠的医生助手,持续推动我国口腔医学服务体系向更高效与高质量的方向迈进。

基金项目

浙江省“领雁”计划项目/省重点研发项目(2025C02102)。

参考文献

- [1] Liu, M., Okuhara, T., Huang, W., Ogihara, A., Nagao, H.S., Okada, H., *et al.* (2025) Large Language Models in Dental Licensing Examinations: Systematic Review and Meta-Analysis. *International Dental Journal*, **75**, 213-222. <https://doi.org/10.1016/j.identj.2024.10.014>
- [2] Meskó, B. (2023) The Impact of Multimodal Large Language Models on Health Care's Future. *Journal of Medical Internet Research*, **25**, e52865. <https://doi.org/10.2196/52865>
- [3] 邓旭亮, 徐明明, 杜宸临, 等. 人工智能驱动口腔医学: 临床、科研、教学与管理的创新探索[J]. 北京大学学报(医学版), 2025, 57(5): 821-826.
- [4] Ojha, M. and Jawarker, R. (2025) Chat Generative Pretrained Transformer: The Present Scenario in Dentistry and What Lies Beyond. *SRM Journal of Research in Dental Sciences*, **16**, 243-246. https://doi.org/10.4103/srmjrd.srmjrd.101_25
- [5] Ronsivalle, V., Santonocito, S., Cammarata, U., Lo Muzio, E. and Cicciù, M. (2025) Current Applications of Chatbots Powered by Large Language Models in Oral and Maxillofacial Surgery: A Systematic Review. *Dentistry Journal*, **13**, Article 261. <https://doi.org/10.3390/dj13060261>
- [6] Sahoo, H.S., Nivedha, R., Nirubama, R. and Kamaleswar, Y. (2025) Performance and Potential of Large Language Models in Restorative Dentistry, Endodontic Diagnosis, and Education: A Systematic Review. *Saudi Endodontic Journal*, **16**, 1-11. https://doi.org/10.4103/sej.sej_121_25
- [7] Abdul, N.S., Shivakumar, G.C., Sangappa, S.B., Di Blasio, M., Crimi, S., Cicciù, M., *et al.* (2024) Applications of Artificial Intelligence in the Field of Oral and Maxillofacial Pathology: A Systematic Review and Meta-Analysis. *BMC Oral Health*, **24**, Article No. 122. <https://doi.org/10.1186/s12903-023-03533-7>
- [8] Tassoker, M. (2025) Exploring ChatGPT's Potential in Diagnosing Oral and Maxillofacial Pathologies: A Study of 123 Challenging Cases. *BMC Oral Health*, **25**, Article No. 1187. <https://doi.org/10.1186/s12903-025-06444-x>
- [9] Alvarez-Silberberg, V.I., Alvarez-Silberberg, C.P., Galletti, C., Flores-Fraile, J., Galletti, C., Ramirez, V., *et al.* (2026) Comparative Analysis of Large Language Models as Decision Support Tools in Oral Pathology. *Scientific Reports*, **16**, Article No. 11272. <https://doi.org/10.1038/s41598-026-41533-z>
- [10] Mohammad-Rahimi, H., Khoury, Z.H., Alamdari, M.I., Rokhshad, R., Motie, P., Parsa, A., *et al.* (2024) Performance of AI Chatbots on Controversial Topics in Oral Medicine, Pathology, and Radiology. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, **137**, 508-514. <https://doi.org/10.1016/j.oooo.2024.01.015>
- [11] Aşar, E.M., İpek, İ. and Bilge, K. (2025) Customized GPT-4V(Ision) for Radiographic Diagnosis: Can Large Language Model Detect Supernumerary Teeth? *BMC Oral Health*, **25**, Article No. 756. <https://doi.org/10.1186/s12903-025-06163-3>
- [12] Kim, K. and Kim, B.C. (2025) Diagnostic Performance of Large Language Models in Multimodal Analysis of Radiolucent Jaw Lesions. *International Dental Journal*, **75**, Article 103910. <https://doi.org/10.1016/j.identj.2025.103910>
- [13] Surana, P., P. Ostwal, P., Vishal Dev, S., Tiwari, J., Charan Yadav, K.S. and Renuka, G. (2024) Role of ChatGPT in

- Dentistry: A Review. *Research Journal of Pharmacy and Technology*, **17**, 3489-3491. <https://doi.org/10.52711/0974-360x.2024.00545>
- [14] Makrygiannakis, M.A., Giannakopoulos, K. and Kaklamanos, E.G. (2025) Evidence-Based Potential of Generative Artificial Intelligence Large Language Models in Orthodontics: A Comparative Study of ChatGPT, Google Bard, and Microsoft Bing. *European Journal of Orthodontics*, **48**, cjae017. <https://doi.org/10.1093/ejo/cjae017>
- [15] Rewthamrongsris, P., Burapachep, J., Trachoo, V. and Pornaveetus, T. (2025) Accuracy of Large Language Models for Infective Endocarditis Prophylaxis in Dental Procedures. *International Dental Journal*, **75**, 206-212. <https://doi.org/10.1016/j.identj.2024.09.033>
- [16] Tokgöz Kaplan, T. and Cankar, M. (2025) Evidence-Based Potential of Generative Artificial Intelligence Large Language Models on Dental Avulsion: ChatGPT versus Gemini. *Dental Traumatology*, **41**, 178-186. <https://doi.org/10.1111/edt.12999>
- [17] Camargo, E.S., Quadras, I.C.C., Garanhani, R.R., de Araujo, C.M. and Stuginski-Barbosa, J. (2025) A Comparative Analysis of Three Large Language Models on Bruxism Knowledge. *Journal of Oral Rehabilitation*, **52**, 896-903. <https://doi.org/10.1111/joor.13948>
- [18] Zhang, R., Pan, Y., Liu, Y., Deng, Y. and Pow, E.H.N. (2025) Leveraging Large Language Models for Patient Instructions in Dentistry—A Systematic Review and Meta-Analysis. *Journal of Prosthodontics*. <https://doi.org/10.1111/jopr.70068>
- [19] Helvacioğlu-Yigit, D., Demirtürk, H., Ali, K., Tamimi, D., Koenig, L. and Almashraqi, A. (2025) Evaluating Artificial Intelligence Chatbots for Patient Education in Oral and Maxillofacial Radiology. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, **139**, 750-759. <https://doi.org/10.1016/j.oooo.2025.01.001>
- [20] Çitir, M. (2025) ChatGPT and Oral Cancer: A Study on Informational Reliability. *BMC Oral Health*, **25**, Article No. 86. <https://doi.org/10.1186/s12903-025-05479-4>
- [21] Aguiar de Sousa, R., Costa, S.M., Almeida Figueiredo, P.H., Camargos, C.R., Ribeiro, B.C. and Alves e Silva, M.R.M. (2024) Is ChatGPT a Reliable Source of Scientific Information Regarding Third-Molar Surgery? *The Journal of the American Dental Association*, **155**, 227-232.e6. <https://doi.org/10.1016/j.adaj.2023.11.004>
- [22] Alhaidry, H.M., Fatani, B., Alrayes, J.O., Almana, A.M. and Alhaed, N.K. (2023) ChatGPT in Dentistry: A Comprehensive Review. *Cureus*, **15**, e38317. <https://doi.org/10.7759/cureus.38317>
- [23] Nguyen, V.A., Vuong, T.Q.T. and Nguyen, V.H. (2025) Benchmarking Large-Language-Model Vision Capabilities in Oral and Maxillofacial Anatomy: A Cross-Sectional Study. *PLOS One*, **20**, e0335775. <https://doi.org/10.1371/journal.pone.0335775>
- [24] Claman, D. and Sezgin, E. (2024) Artificial Intelligence in Dental Education: Opportunities and Challenges of Large Language Models and Multimodal Foundation Models. *JMIR Medical Education*, **10**, e52346-e52346. <https://doi.org/10.2196/52346>
- [25] Thorat, V.A., Rao, P., Joshi, N., Talreja, P. and Shetty, A. (2024) The Role of Chatbot GPT Technology in Undergraduate Dental Education. *Cureus*, **16**, e54193. <https://doi.org/10.7759/cureus.54193>
- [26] Benítez, T.M., Xu, Y., Boudreau, J.D., Kow, A.W.C., Bello, F., Van Phuoc, L., et al. (2024) Harnessing the Potential of Large Language Models in Medical Education: Promise and Pitfalls. *Journal of the American Medical Informatics Association*, **31**, 776-783. <https://doi.org/10.1093/jamia/ocad252>
- [27] Künzle, P. and Paris, S. (2024) Performance of Large Language Artificial Intelligence Models on Solving Restorative Dentistry and Endodontics Student Assessments. *Clinical Oral Investigations*, **28**, Article No. 575. <https://doi.org/10.1007/s00784-024-05968-w>
- [28] Al-Moghrabi, D., Abu Arqub, S., Maroulakos, M.P., Pandis, N. and Fleming, P.S. (2024) Can ChatGPT Identify Predatory Biomedical and Dental Journals? A Cross-Sectional Content Analysis. *Journal of Dentistry*, **142**, Article 104840. <https://doi.org/10.1016/j.ident.2024.104840>
- [29] Freire, Y., Santamaría Laorden, A., Orejas Pérez, J., Gómez Sánchez, M., Díaz-Flores García, V. and Suárez, A. (2024) ChatGPT Performance in Prosthodontics: Assessment of Accuracy and Repeatability in Answer Generation. *The Journal of Prosthetic Dentistry*, **131**, 659.e1-659.e6. <https://doi.org/10.1016/j.prosdent.2024.01.018>
- [30] Ohta, K. and Ohta, S. (2023) The Performance of GPT-3.5, GPT-4, and Bard on the Japanese National Dentist Examination: A Comparison Study. *Cureus*, **15**, e50369. <https://doi.org/10.7759/cureus.50369>
- [31] Buldur, M. and Sezer, B. (2024) Evaluating the Accuracy of Chat Generative Pre-Trained Transformer Version 4 (ChatGPT-4) Responses to United States Food and Drug Administration (FDA) Frequently Asked Questions about Dental Amalgam. *BMC Oral Health*, **24**, Article No. 605. <https://doi.org/10.1186/s12903-024-04358-8>
- [32] Hager, P., Jungmann, F., Holland, R., Bhagat, K., Hubrecht, I., Knauer, M., et al. (2024) Evaluation and Mitigation of the Limitations of Large Language Models in Clinical Decision-Making. *Nature Medicine*, **30**, 2613-2622.

- <https://doi.org/10.1038/s41591-024-03097-1>
- [33] Zhu, G., Zhang, X. and Chen, C. (2025) Assessing and Enhancing the Reliability of Chinese Large Language Models in Dental Implantology. *BMC Oral Health*, **25**, Article No. 1242. <https://doi.org/10.1186/s12903-025-06648-1>
- [34] Hooshiar, M.H. (2025) Artificial Intelligence Reliability in Implant Dentistry: A Comparative Analysis of Clinical Accuracy and Hallucination Patterns across Multiple Language Models. *The Journal of Prosthetic Dentistry*, Articles in Press. <https://doi.org/10.1016/j.prosdent.2025.11.005>
- [35] Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., *et al.* (2023) Opinion Paper: “So What If ChatGPT Wrote It?” Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy. *International Journal of Information Management*, **71**, Article 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- [36] Baskar, S., Bhuvana, R. and Hemalatha, R.J. (2024) Introduction to Smart Hospital. Wiley Online Library.
- [37] Gao, Y., Luo, W., Wang, C., Ahmad, N.S., Wang, X. and Goh, P. (2026) A Privacy-Preserving Multi-User Retrieval System for Multimodal Artificial Intelligence. *Scientific Reports*, **16**, Article No. 10348. <https://doi.org/10.1038/s41598-026-40734-w>
- [38] Kumar, J., Almustafa, K.M., Madanat, R., Sharma, A.K., Sutcu, M. and Katrib, J. (2025) Privacy-Aware and Interpretable Deep Learning Framework for Dental Caries Classification. *Intelligence-Based Medicine*, **12**, Article 100294. <https://doi.org/10.1016/j.ibmed.2025.100294>
- [39] Sallam, M. (2023) ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*, **11**, Article 887. <https://doi.org/10.3390/healthcare11060887>
- [40] Consorti, G., Catarzi, L., Frosolini, A., Vaira, L.A., Committeri, U. and Cirignaco, G. (2026) Artificial Intelligence in Oral and Maxillofacial Surgery: A Scoping Review of Clinical Applications, Ethical Challenges, and Legal Considerations. *International Journal of Oral and Maxillofacial Surgery*, Articles in Press. <https://doi.org/10.1016/j.ijom.2026.02.020>
- [41] Ahn, C. (2023) Exploring ChatGPT for Information of Cardiopulmonary Resuscitation. *Resuscitation*, **185**, Article 109729. <https://doi.org/10.1016/j.resuscitation.2023.109729>
- [42] AlSaad, R., Abd-Alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M., Damseh, R., *et al.* (2024) Multimodal Large Language Models in Health Care: Applications, Challenges, and Future Outlook. *Journal of Medical Internet Research*, **26**, e59505. <https://doi.org/10.2196/59505>
- [43] Huang, Y., Yuan, Q., Sheng, X., Yang, Z., Wu, H., Chen, P., Yang, Y., Li, L. and Lin, W. (2024) AesBench: An Expert Benchmark for Multi-Modal Large Language Models on Image Aesthetics Perception. arXiv:2401.08276.
- [44] Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., *et al.* (2024) Explainability for Large Language Models: A Survey. *ACM Transactions on Intelligent Systems and Technology*, **15**, 1-38. <https://doi.org/10.1145/3639372>
- [45] 孙运梁, 王璨. 医疗人工智能介入过失诊疗行为的结果归属研究[J]. 江西师范大学学报(哲学社会科学版), 2025, 58(5): 137-148.