THE JOURNAL OF PROSTHETIC DENTISTRY

RESEARCH AND EDUCATION

Can artificial intelligence (AI)-based software programs generate accurate clinical dictation?

Sohil A. Kazim, BDS, MDS,[a] Charles J. Goodacre, DDS, MSD,[b] Joseph Y.K. Kan, DDS, MS,[c] and Gary R. Goldstein, DDS[d]

ABSTRACT

Statement of problem. Clinical dictation in dental practice is time-consuming and prone to quality variability. Although generative artificial intelligence (AI) systems such as Microsoft Copilot, OpenAI ChatGPT, Google Gemini, and OpenEvidence offer potential alternatives for automating and enhancing this process, their ability to produce clinically reliable and contextually appropriate dictations has not been evaluated.

Purpose. The purpose of this evaluation study was to assess the performance of Copilot, 2 versions of ChatGPT (v4 and 5), Gemini (v2.5), and OpenEvidence in generating accurate, complete, and clinically relevant dental dictations based on varying prompt structures.

Material and methods. Two different representative clinical scenarios were generated. One described the placement of a single mandibular implant with the dictation focused on the surgical steps, while the second described a prosthodontic procedure and the steps required to place a crown on a single maxillary implant. Each scenario was entered into Copilot, 2 different versions of ChatGPT, Gemini, and OpenEvidence using a structured prompt format (SP). All inputs were conducted in a clean session. All responses from Copilot, both included versions of ChatGPT, Gemini, and OpenEvidence software programs, were compared for both scenarios. Minimum specific prompts were also generated for the same clinical scenarios and tested to determine the least amount of clinician input required to produce acceptable dictations. Overall responses were compared with the more detailed structured prompts to assess differences in output quality.

Results. Across both surgical and restorative procedures, all 5 AI software programs (Copilot, ChatGPT-4, ChatGPT-5, Gemini, and OpenEvidence) produced clinically accurate dictations when using the SP format. For the surgical procedure, core steps—including preoperative assessment, anesthesia, incision design, osteotomy preparation, implant placement, closure, and postoperative instructions—were consistent. For the restorative procedure, all software programs documented atraumatic healing abutment removal, healthy peri-implant mucosa, correct crown fit, occlusion, shade matching, and standardized tightening (35 Ncm). No substantial differences in procedural accuracy were identified; discrepancies were primarily stylistic or related to the level of descriptive detail. With minimum specific prompts, the dictations were shorter and less descriptive but still captured the essential procedural steps for both clinical scenarios and were considered clinically acceptable in all included AI software programs.

Conclusions. Microsoft Copilot, both versions of OpenAI ChatGPT (GPT-4 and GPT-5), Google Gemini, and OpenEvidence can effectively assist clinicians in generating clinical dental procedure dictations, particularly when guided with precise and structured prompts. Minimum specific prompts produced shorter but clinically acceptable outputs that may be particularly useful in busy clinical settings. While not a substitute for clinician oversight, the AI technology promises to improve efficiency and documentation consistency in dental practice. (J Prosthet Dent 2026;135:596.e1-e7)

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
The authors declare no conflicts of interest.

[a]Fellow, Advanced Education Program in Implant Dentistry, School of Dentistry, Loma Linda University, Loma Linda, Calif; and Adjunct Professor, Department of Restorative Dentistry, Advanced Specialty Education Program in Prosthodontics, Rutgers School of Dental Medicine, Rutgers University, Newark, NJ.
[b]Distinguished Professor, Advanced Education Program in Implant Dentistry, School of Dentistry, Loma Linda University, Loma Linda, Calif.
[c]Professor, Advanced Education Program in Implant Dentistry, School of Dentistry, Loma Linda University, Loma Linda, Calif.
[d]Professor Emeritus, Department of Prosthodontics, College of Dentistry, New York University, New York, NY.

Clinical Implications

Artificial intelligence–based software programs such as Copilot, ChatGPT, Gemini, and OpenEvidence can reliably generate clinically acceptable dental dictations when provided with structured prompts. These tools may streamline documentation, reduce variability, and enhance efficiency in clinical practice. However, clinician oversight remains essential to verify accuracy, ensure medicolegal compliance, and tailor notes to patient-specific contexts. The integration of AI dictation into dental workflows may also serve as an educational aid for trainees by reinforcing procedural sequencing and terminology.

Clinical documentation is an essential component of contemporary dental practice, and dictation is key to maintaining accurate records, facilitating interdisciplinary communication, and supporting medicolegal requirements.[1] However, clinicians often face challenges related to time constraints, inconsistent narrative quality, and the cognitive load associated with documentation.[2,3]

The emergence of large language models (LLMs) introduces the possibility of using artificial intelligence (AI) to generate clinical dictation. Unlike traditional speech-to-text transcription tools, LLMs offer context-aware, semantically organized, and grammatically coherent content.[4,5]

Dental-related digital software programs that can assist in generating clinical dictations, such as Bola AI, DentScribe.ai, Denti.AI Scribe, Denota.ai, Dentistry Dashboard's AI Dental Notes Creator, Dragon Medical One, Kiroku, Overjet, and TapNote, are available. However, OpenAI ChatGPT was the most popular AI software program as of August 2025. Seven hundred million users interact with ChatGPT weekly worldwide, with Google Gemini, with about 450 million monthly users, in second place. ChatGPT was unveiled in November 2022 as a digital information resource to address inquiries on a variety of scientific subjects.[6] Its vast database, which has been compiled from books, articles, and websites, imitates statistical patterns of language using an LLM.[7] The architecture of ChatGPT has been reported to create responses based on the text input by using a neural network.[8] OpenAI launched ChatGPT-4 in March 2023, an improved version intended to more efficiently respond to complex queries.[9,10] More recently, ChatGPT-5, which requires a fee, was released in August 2025. Microsoft introduced Copilot in 2023 and soon after was embedded in Microsoft 365 apps, including Word, Excel, and PowerPoint, and then in Edge on the browser sidebar. In 2024, it became available on iOS and Android. Gemini was launched in December of 2023 as an advanced replacement of Bard, which had been launched earlier that year. Its initial version 1 has been improved and version 2.5 was released in the spring of 2025. As with ChatGPT, there are tiered options.

Because of the popularity of ChatGPT, Gemini, and Copilot, the present comparison study was conducted to evaluate their reliability and effectiveness in providing high-quality and complete dental dictations. In addition to Microsoft Copilot, OpenAI ChatGPT, and Google Gemini, OpenEvidence, another recently introduced LLM, was launched in 2023 in partnership with the Mayo Clinic, with contractual collaborations through the New England Journal of Medicine and the Journal of the American Medical Association. OpenEvidence is unique among current LLMs because it incorporates a hierarchy of clinical evidence when generating content.[11] In addition to describing procedural steps, it can summarize long-term outcomes, complications, and prognosis.[11,12] This evidence-based functionality provided an added layer of clinical relevance not consistently found in other AI models.[11] The research hypothesis was that Microsoft Copilot, OpenAI ChatGPT (GPT-4 and GPT-5), Google Gemini, and OpenEvidence could generate accurate and complete dictations when using a structured prompt format (SP) clinical scenario.

MATERIAL AND METHODS

Two clinical scenarios were created using a structured prompt (SP) format. Scenario 1 described the placement of a single mandibular implant, with the dictation focused on the surgical steps, while scenario 2 described a prosthodontic procedure and the steps required to place a crown on a single maxillary implant. The SP format was organized with specific detailed input. Each scenario was input into Microsoft Copilot, OpenAI ChatGPT (GPT-4 and GPT-5), Google Gemini (V 2.5), and OpenEvidence in a clean session.

Scenario 1: Generate a detailed clinical surgical dictation for dental implant placement in a healed site mandibular left first molar, including pre-operative evaluation, local anesthesia type, incision and flap design, implant osteotomy sequence, implant system (NobelActive, TiUltra, diameter of 5.0 mm, length of 10 mm), torque value, closure technique, and post-operative instructions.

Scenario 2: Generate a detailed clinical dictation for placement of an implant-supported fixed screw-retained single zirconia restoration for a maxillary right first molar NobelActive implant. Include implant system and diameter (5.0 mm), healing abutment removal, evaluation of peri-implant mucosa, verification of crown fit (contacts, Ti-base abutment fit, emergence profile), evaluation of occlusion, evaluation of shade and contour, abutment screw torque value, material used to close the screw access channel, and postinsertion instructions.

The responses from Copilot, both versions of ChatGPT, Gemini, and OpenEvidence software programs were compared for both SP scenarios. To further assist clinicians, minimum specific prompts were also developed for the same clinical scenarios to test the level of detail required to generate a clinically acceptable dictation. Each scenario was input into the same included software programs. These simplified prompts represented what a busy practitioner could enter immediately after a procedure while still maintaining accuracy. Scenario 1: Generate a clinical dictation for the placement of a dental implant in the mandibular left first molar. Scenario 2: Generate a clinical dictation for the placement of a screw-retained zirconia restoration on a maxillary right first molar dental implant. The overall responses from minimum specific prompts and structured prompts were compared for both clinical scenarios.

RESULTS

Dictations generated by Copilot, both included versions of ChatGPT (GPT-4 and GPT-5), Gemini, and OpenEvidence, were deemed clinically acceptable for both scenarios. The results from Copilot, both versions of ChatGPT, Gemini, and OpenEvidence were very similar to each other, showing only minor differences across all responses to the clinical scenarios. As a result, a list of items that were compared between all 5 AI software programs was generated for each clinical scenario to show the variations (Tables 1 and 2).

In scenario 1 (surgical dictation), the comparison of implant surgical dictations generated by Copilot, ChatGPT-4, ChatGPT-5, Gemini, and OpenEvidence demonstrated differences in descriptive detail, terminology, and procedural sequencing, despite consistency in core surgical parameters. All 5 platforms correctly identified the implant system (NobelActive TiUltra, 5.0×10 mm; Nobel Biocare Services AG) and site (mandibular left first molar) and documented successful completion without intraoperative complications.

In the preoperative evaluation, ChatGPT-4 provided the most comprehensive detail. Copilot and ChatGPT-5 offered concise assessments, while Gemini and OpenEvidence additionally noted systemic health.

Anesthesia protocols varied in specificity. Copilot, ChatGPT-5, and OpenEvidence specified lidocaine 2% with precise carpule counts and multiple blocks and/or infiltrations. ChatGPT-4 recorded the use of both lidocaine and articaine, whereas Gemini specified 4% articaine without specifying quantity.

In incision and flap design, Copilot, ChatGPT-4, ChatGPT-5, and OpenEvidence uniformly described a mid-crestal incision with vertical releases and a full-thickness flap. Gemini differed, using a crestal incision combined with buccal sulcular and lingual releases.

For the osteotomy sequence, all AI software programs reported sequential drilling to a final 5.0-mm diameter under irrigation. Copilot included all intermediate diameters from pilot drill to final width. ChatGPT-4 and ChatGPT-5 omitted the 2.0-mm step, Gemini and OpenEvidence did not specify intermediate diameters. ChatGPT-5 uniquely documented intraoperative radiographic verification. Implant placement was uniformly described as the insertion of a NobelActive TiUltra implant to 45 Ncm with cover screw placement.

Closure techniques differed in suture material selection. Copilot and ChatGPT-5 used 4-0 PTFE in interrupted and mattress patterns, ChatGPT-4 used 4-0 chromic gut interrupted sutures, and Gemini used 4-0 Vicryl interrupted sutures. OpenEvidence used 4-0 monofilament sutures without specifying the exact suture material.

Postoperative instructions varied in the duration of dietary modification and activity restrictions. Copilot and ChatGPT-5 advised avoiding mastication for 7 to 10 days. ChatGPT-4 recommended a 3- to 5-day soft diet and avoidance of straw use, whereas Gemini and OpenEvidence prescribed a 7-day soft diet with a 48-hour activity limitation. Antibiotic and chlorhexidine regimens were consistently included, with follow-up scheduled between 7 and 10 days.

In the outcome section, all platforms recorded uncomplicated procedures and stable postoperative conditions. ChatGPT-4 uniquely noted minimal blood loss and provided a planned osseointegration interval of 3 to 4 months.

In scenario 2 (implant crown placement dictation), across all included platforms, core procedural steps for the placement of an implant-supported fixed screw-retained single zirconia crown for a maxillary right first molar implant were consistent, with all noting atraumatic healing abutment removal, stable peri-implant mucosa, proper crown fit, occlusion, shade matching, torque values, and screw access channel closure. Overall, no substantial differences in procedural accuracy or core restorative steps were observed between all included AI software programs. Variations were primarily in descriptive style, specificity of clinical details, and postinsertion follow-up recommendations.

Moreover, when using minimum specific prompts, all included software programs produced dictations that were shorter and less descriptive than the SP but still clinically acceptable. SP consistently yielded more comprehensive detail; however, the minimum specific

Table 1. Items comparing responses from Copilot, both versions of ChatGPT (GPT-4 and GPT-5), Gemini, and OpenEvidence software programs for clinical scenario 1 in SP format (surgical implant placement)

Section	Copilot	ChatGPT-4	ChatGPT-5	Gemini	OpenEvidence
Title	Clinical Surgical Dictation - CP	Surgical Placement of Dental Implant in Mandibular Left First Molar	Clinical Surgical Dictation	Gemini Dictation	Mandibular Molar Implant Dictation
Patient Info	Patient Name, DOB, MRN, Date of Procedure	Patient Name, DOB, Date of Surgery, Tooth Site (mandibular left first molar)	Patient Name, DOB, Date of Procedure	Patient Name, DOB, Date of Surgery	Not included
Procedure	NobelActive TiUltra 5.0x10 mm	NobelActive TiUltra 5.0x10 mm	NobelActive TiUltra 5.0x10 mm	NobelActive TiUltra 5.0x10 mm	NobelActive TiUltra 5.0×10 mm
Pre-Operative Evaluation	Healed site, adequate tissue, CBCT confirms bone volume, no vital structure proximity	Well-healed ridge, CBCT shows 13 mm height to IAN, 7 mm width, bone quality D2, no pathology, SAC classification straightforward	Healed ridge, CBCT confirms adequate bone, no pathology, safe from IAN	Healed site, adequate bone volume, CBCT confirms safe from IAN, non-smoker, healthy	Healed ridge, CBCT confirms adequate bone; notes systemic health and absence of pathology
Anesthesia	2 carpules lidocaine 2%, IAN block+buccal	IAN block+buccal and lingual infiltration - lidocaine & articaine	2 carpules lidocaine 2%, IAN block+long buccal+lingual	4% articaine, IAN block+buccal	Lidocaine 2% (2 carpules) IAN block with buccal and lingual infiltration
Incision and Flap Design	Mid-crestal with 2 vertical releases, full-thickness flap	Mid-crestal with 2 vertical releases, full-thickness flap	Mid-crestal with 2 vertical releases, full-thickness flap	Crestal incision+buccal sulcular+lingual release, full-thickness flap	Mid-crestal incision with vertical releases; full-thickness flap
Osteotomy Sequence	Pilot drill → 2.0 mm → 3.5 mm → 4.3 mm → final 5.0 mm, saline irrigation	Pilot drill → 3.5 mm → 4.3 mm → final 5.0 mm, 10 mm depth, saline irrigation	Pilot drill → 3.5 mm → 4.3 mm → final 5.0 mm, saline irrigation, radiographic check	Pilot drill → sequential to final 5.0 mm, saline irrigation	Pilot drill → sequential to final 5.0 mm, saline irrigation
Implant Placement	NobelActive TiUltra, torque 45 Ncm, cover screw placed	NobelActive TiUltra, torque 45 Ncm, cover screw placed	NobelActive TiUltra, torque 45 Ncm, cover screw placed	NobelActive TiUltra, torque 45 Ncm, cover screw placed	NobelActive TiUltra, torque 45 Ncm, cover screw placed
Closure Technique	4-0 PTFE interrupted+mattress sutures	4-0 chromic gut interrupted sutures	4-0 PTFE interrupted+mattress sutures	4-0 Vicryl interrupted sutures	4-0 monofilament sutures (interrupted pattern); material not specified
Post-Operative Instructions	Avoid chewing 7-10 d, antibiotics, ibuprofen, chlorhexidine, ice, follow-up 1 wk	Soft diet 3-5 d, ice, no straws, chlorhexidine, antibiotics, pain meds, follow-up 7-10 d	Avoid chewing 7-10 d, antibiotics, ibuprofen, chlorhexidine, follow-up 7-10 d	Soft diet 7 d, antibiotics, ibuprofen, chlorhexidine, avoid activity 48 h, follow-up 1 wk	7-d soft diet; antibiotics and chlorhexidine rinse prescribed; 48-h activity restriction; follow-up in 7-10 d
Notes and Outcome	No complications, tolerated well	No complications, minimal blood loss, osseointegration planned 3-4 mo	No complications, stable discharge	No complications	No complications, stable discharge

Table 2. Items comparing responses from Copilot, both versions of ChatGPT (GPT-4 and GPT-5), Gemini, and OpenEvidence software programs for clinical scenario 2 in SP format (implant crown placement)

Section	Copilot	ChatGPT-4	ChatGPT-5	Gemini	OpenEvidence
Title	Placement of Implant-Supported Fixed Screw-Retained Single Zirconia Crown—maxillary right first molar implant	Clinical Procedure Note: Implant-Supported Restoration Placement—Site (maxillary right first molar)	Clinical Dictation—Implant-Supported Fixed Screw-Retained Single Zirconia Restoration	Delivery of Fixed Screw-Retained Zirconia Crown, maxillary right first molar implant	Implant Restoration Placement
Healing Abutment Removal	Healing abutment removed under sterile conditions	Atraumatic removal with driver	Removed with appropriate driver	Carefully unscrewed and removed	Removed under sterile conditions with driver
Peri-Implant Mucosa Evaluation	Healthy, pink, stippled, no inflammation	Healthy, no inflammation, adequate keratinization	Healthy, adequate keratinized gingiva, no inflammation	Firm, pink, healthy tissue, papillae present	Healthy, keratinized tissue; no inflammation
Verification of Crown Fit (Contacts, Ti-base, Emergence Profile)	Contacts ideal; Ti-base fit passive; emergence profile harmonious	Contacts tight but correct; Ti-base passive fit; emergence profile natural	Contacts correct; Ti-base fully seated; emergence profile optimal	Contacts solid but not excessive; Ti-base seated; emergence profile well-contoured	Passive Ti-base fit; proper contacts; harmonious emergence profile
Occlusion Evaluation	Light centric contacts, no interferences	Balanced occlusion after adjustments	Adjusted high spots for balanced occlusion	Minor adjustments; out of occlusion in excursions	Light centric contacts, no interferences
Shade and Contour Evaluation	Highly esthetic, shade match to adjacent dentition	Shade and contour match adjacent teeth	Shade and contour match adjacent dentition	Shade A1, excellent match, patient satisfied	Shade and contour match adjacent dentition
Abutment Screw Torque Value	35 Ncm	35 Ncm	35 Ncm	35 Ncm	35 Ncm
Screw Access Channel Closure	Teflon tape+composite resin	Teflon tape+light-cured composite resin	Teflon tape+light-cured composite resin	PTFE tape+opaque composite resin	PTFE tape+light-cured composite resin
Post-Insertion Instructions	Avoid hard foods 24 h; maintain oral hygiene; follow-up in 2 wk	Avoid chewing 24 h; hygiene with interdental aids; 6-mo maintenance; follow-up 2 wk	Avoid hard and sticky foods 24 h; hygiene instructions; follow-up 2 wk	Avoid hard and sticky foods 24 h; maintain hygiene; follow-up 1 mo	Avoid hard or sticky foods for 24 h; maintain hygiene; follow-up in 2 wk

prompts were sufficient to generate usable clinical dictations when time efficiency is a priority.

In the minimum specific prompt for the surgical implant placement scenario, the dictations produced by all included software programs were briefer than those generated from the SP. Core procedural steps such as anesthesia, implant placement with torque, closure, and postoperative instructions were consistently documented. However, some details included in the SP dictations—such as preoperative assessment, incision design, osteotomy sequence, and suture material—were generally absent (Table 3).

In the minimum specific prompt for the implant crown placement scenario, the outputs captured key restorative steps, including healing abutment removal, crown fit, occlusion, torque, screw access closure, and postinsertion instructions. Compared with the SP responses, the dictations omitted peri-implant mucosal description, shade and contour evaluation, and specific closure materials for the screw-access channel (Table 4).

DISCUSSION

This evaluation of AI-generated clinical dictations provided evidence that Copilot, ChatGPT (GPT-4 and GPT-5), Gemini, and OpenEvidence could produce high-quality clinical dictations suitable for dental documentation when using proper and detailed structured prompt input. A list of items that compared the responses from all included AI software programs was created for each clinical scenario (Tables 1, 2). Therefore, the research hypothesis that Copilot, ChatGPT (GPT-4 and GPT-5), Gemini, and OpenEvidence could generate accurate and complete dictations when using a structured prompt format (SP) clinical scenario was accepted.

AI-generated dictations are not immune from occasional omissions, assumptions, or overly generalized language. Therefore, AI tools must be viewed as clinical extenders, not decision-makers or documentation substitutes. Integration into clinical workflow must include human validation and editing.

A significant issue with Copilot was that it had access to the user's "context clues," which allowed Microsoft to save browsing activity, including history, usage, favorites, content, web content, and other browsing data. This allowed Microsoft to personalize ads, search, shopping, and news. It would not download references unless "context clues" had been toggled. ChatGPT and Gemini did not have access to the user's computer or browsing history. They had access only to files uploaded to them.

Having progress notes that are more extensive than those typically placed in patient charts has an additional value. They can serve as good educational tools to teach students the procedures and sequence involved in specific

Table 3. Items comparing overall responses from structured prompts and minimum specific prompts for clinical scenario 1 (surgical implant placement)

Section	Structured Prompt (SP)	Minimum Specific Prompt
Prompt content	Detailed: pre-operative evaluation, anesthesia, incision, flap design, osteotomy sequence, implant system, dimensions, torque, closure, post-op instructions	Concise: anesthesia, implant placement, torque, closure, post-op instructions
Pre-operative evaluation	Ridge dimensions, CBCT findings, bone quality, systemic health	Usually omitted; assumed adequate and straightforward procedure
Anesthesia	Specified type, carpule count, injection sites	Mentioned briefly (such as, local anesthesia given)
Incision and Flap design	Mid-crestal incision ±vertical releases, flap thickness specified	General mention (such as, flap raised)
Osteotomy and Implant placement	Sequential drilling, irrigation, implant system, dimensions, torque values	Implant placed in mandibular left first molar with adequate torque, no drill sequence
Closure	Specific sutures and techniques described	Closure noted, no material specified
Post-op instructions	Detailed: antibiotics, chlorhexidine, diet, activity, follow-up	Mentioned briefly (such as, post-op instructions provided)
Overall output	Comprehensive surgical dictation	Shorter but clinically acceptable dictation for busy clinicians

Table 4. Items comparing overall responses from structured prompts and minimum specific prompts for clinical scenario 2 (implant crown placement)

Section	Structured Prompt (SP)	Minimum Specific Prompt
Prompt content	Detailed: healing abutment removal, evaluation of peri-implant mucosa, crown fit (contacts, Ti-base, emergence), occlusion, shade, torque, screw access closure, post-insertion instructions	Concise: healing abutment removal, evaluation of crown fit, occlusion, torque, screw access closure, post-insertion instructions
Healing abutment removal	Atraumatic, sterile technique, specific instrument documented	Brief mention (such as, healing abutment removed)
Peri-implant mucosa	Documented health, keratinization, papillae	Omitted
Crown fit	Detailed: contacts, Ti-base seating, emergence profile	Brief mention (such as, crown fitted and seated with proper contacts)
Occlusion	Adjustments described, excursions recorded	Brief mention (such as, occlusion checked and adjusted)
Shade and Contour	Esthetics documented; sometimes with shade code	Mentioned briefly (such as, shade and contour were checked)
Torque	Reported as 35 Ncm across all AI software programs	Mentioned as “abutment screw torqued”
Screw access closure	Materials described (PTFE+composite)	General: “screw access closed”
Post-insertion instructions	Specific hygiene, diet, follow-up interval	Brief: “instructions provided”
Overall output	Comprehensive restorative dictation	Concise, efficient dictation suitable for busy practice

treatments. Also, they can be used to remind clinicians of information they may want to include in their notes.

The inclusion of minimum specific prompt scenarios demonstrated that even concise clinician input could generate acceptable dictations. While structured prompts provided the most detail and completeness, minimum specific prompts might be particularly valuable in high-volume practice settings where efficiency is essential (Tables 3, 4).

Furthermore, the addition of OpenEvidence highlights another dimension in AI dictation. Unlike the other evaluated platforms, OpenEvidence emphasized evidence-based outcomes and prognosis, drawing upon a structured hierarchy of medical literature. This functionality could enrich charting by not only documenting procedural steps but also contextualizing long-term expectations and potential complications. As such, OpenEvidence might serve as a complementary adjunct to existing LLMs.

Limitations of this evaluation study of AI-generated dictations included that there were only 2 scenarios, both simulated. Only 4 clinicians evaluated the scenarios, and the study focused on English language output only. The evaluation did not include integration into actual electronic health record (EHR) systems.

CONCLUSIONS

Based on the findings of this evaluation study, the following conclusions were drawn:

1. Microsoft Copilot, both included versions of OpenAI ChatGPT (GPT-4, GPT-5), Google Gemini, and OpenEvidence, when appropriately prompted, could generate coherent and clinically acceptable dental dictations.
2. All included AI software programs produced nearly identical results when responding to both clinical scenarios in SP format.
3. SP consistently produced the most detailed and comprehensive results, but minimum specific prompts were sufficient to generate acceptable and usable clinical dictations.
4. AI-driven dictation tools might serve as valuable adjuncts in clinical settings, provided proper oversight is maintained.

REFERENCES

1. Garcia AC, David GC, Chand D. Understanding the work of medical transcriptionists in the production of medical records. *Health Inform J.* 2010;16:87–100.
2. Farri O, Bak P, Lelie D, Schwartz J. Effects of time constraints on clinician-computer interaction. *J Biomed Inform.* 2013;46:594–600.
3. Moy AJ, Campbell SR, Swartz R, et al. Measurement of clinical documentation burden among physicians and nurses: A scoping review. *J Am Med Inform Assoc.* 2021;28:998–1006.
4. Sallam M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare.* 2023;11:887.
5. Semrl N, Feigl S, Taumberger N, et al. AI language models in human reproduction research: Exploring ChatGPT's potential to assist academic writing. *Hum Reprod.* 2023;38:2281–2288.
6. Macdonald C, Adeloye D, Sheikh A, Rudan I. Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. *J Glob Health.* 2023;13:01003.
7. Stokel-Walker C. ChatGPT listed as author on research papers: Many scientists disapprove. *Nature.* 2023;613:620–621.
8. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? *Crit Care.* 2023;27:75.
9. Abdullahi T, Singh R, Eickhoff C. Learning to make rare and complex diagnoses with generative AI assistance: Qualitative study of popular large language models. *JMIR Med Educ.* 2024;10:e51391.
10. Wang YM, Chen TJ. ChatGPT surges ahead: GPT-4 has arrived in the arena of medical research. *J Chin Med Assoc.* 2023;86:784–785.
11. Hurt RT, Stephenson CR, Gilman EA, et al. The use of an artificial intelligence platform OpenEvidence to augment clinical decision-making for primary care physicians. *J Prim Care Community Health.* 2025;16:21501319251332215.
12. Patel N, Grewal H, Buddhavarapu V, Dhillon G. OpenEvidence: Enhancing medical student clinical rotations with AI but with limitations. *Cureus.* 2025;17:e76867.

Corresponding author:
Dr Sohil A. Kazim
Advanced Education Program in Implant Dentistry
School of Dentistry
Loma Linda University
Loma Linda, CA 92373
Email: sohil.a.k@hotmail.com

https://doi.org/10.1016/j.prosdent.2025.10.051