International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-13, Issue-10; Oct, 2026
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.1310.2



An AI Virtual Receptionist in Dental Practice: A Multi-Site Observational Analysis of 3,532 Patient Calls

Usman Tariq[1]

DentalAssist.ai, United States
Email: usman@dentalassist.ai

Received: 18 Sep 2026,

Received in revised form: 15 Oct 2026,

Accepted: 21 Oct 2026,

Available online: 30 Oct 2026

©2026 The Author(s). Published by AI Publication. This is an open-access article under the CC BY license

(https://creativecommons.org/licenses/by/4.0/).

***Keywords*— AI virtual receptionist, appointment triage, dental practice management, patient scheduling, workflow efficiency.**

***Abstract*—** *This study evaluates how an AI-driven virtual receptionist (DentalAssist.ai) handles incoming patient calls in dental practices and grounds a conceptual framework for AI-driven dental reception in empirical evidence on workflow effects, intent distribution, after-hours coverage and conversational patterns. An observational multi-site design was applied to 3,532 anonymised call records collected from ten North American dental practices over 9.9 months (June 2025 – March 2026), each described by 29 attributes covering timestamps, intent classification, outcome status, booking completion and after-hours flags. Descriptive analyses were complemented by qualitative content coding of anonymised dialogue transcripts (66 extracted, 50 analysable). Across all calls, 67.6% reached a Success outcome, 28.1% were Incomplete and 4.3% were classified as Other. Performance changed substantially across the observation window: Success rose from 26.2% in the first three months (n = 978) to 86.2% in the final three (n = 1,308), booking completion from 3.3% to 32.8%, and abandonment fell from 21.0% to 9.4%. The aggregate figures reported here therefore average over a maturing deployment and should be read alongside that trajectory. Among calls with booking intent (n = 1,830), 38.6% (707) concluded with a confirmed booking, and 27.3% of all calls (n = 964) occurred outside business hours and were received and processed without staff involvement, 66.9% of them reaching a Success outcome. Abandoned calls accounted for 15.1%, of which 94% involved unclear or incomplete intent. Two clinics generated 97% of traffic, with success rates of 63.4% and 93.3% respectively. Qualitative analysis identified a four-stage booking flow, an emergency triage protocol based on capture-and-handover rather than autonomous slot assignment, and human-escalation triggers in 7 of the 50 analysed dialogues. The findings indicate that an AI virtual receptionist in commercial deployment can absorb a measurable share of administrative call volume, including extensive after-hours coverage, while preserving escalation paths, and supply a practice-grounded reference point for future evaluations of conversational AI in dental front-desk operations..*

I. INTRODUCTION

Dental practices operate under intense economic pressure: chair time is the primary revenue source, while clinician minutes are constrained by clinical safety and regulatory requirements. Every unfilled slot, late cancellation, or scheduling error results in lost income and contributes to staff burnout at the reception desk. At the same time, patients expect smooth digital communication such as messengers, online appointment booking, and instant answers to routine questions.

AI-enabled conversational agents are currently used to provide remote care, administrative support and access to

care pathways in larger hospital systems. Rapid reviews and scoping syntheses have described how chatbots have been deployed to support patient education, behavior change, monitoring, and basic administrative responsibilities, such as appointment management and navigation [1, 2]. Broader reviews of healthcare chatbots note possible benefits to efficiency, but also highlight limitations, including safety, explainability and ethics [2, 3]. For dental clinics, which often have limited IT budgets and human resources, the challenge is to distill these broad capabilities into reliable products that integrate with their practice management.

There is a considerable potential for more efficient resource utilization through AI scheduling. A metanarrative review of AI-based patient scheduling in a real-world healthcare system concluded that a range of machine learning and optimization models and objectives have been applied, including reducing no-shows, shortening wait times, and balancing service provider workload [4]. Studies in operations research likewise stress this perspective in relation to healthcare appointment scheduling and capacity planning [5–8], and concepts of clinician supported decision-making, human in the loop, and conversational systems in structured workflows, for example, are reflected in conceptual and empirical studies of person-centered virtual care bots [9]. Together, these trajectories provide the methodological and conceptual background for a dental-specific framework.

Dentistry already represents an active testbed for conversational AI. Research has been done on the performance and readability of general purpose chatbots answering common questions about dental prostheses. Overall, the quality was high, but there was variability across studies [10]. Other studies have compared large language model outputs to those of dentists in terms of informativeness and empathy [11]. In dental education research, AI chatbots have been used as virtual patients to train students in diagnostics and communication. These results were positively evaluated by students, showing high learning gains [12]. These studies indicate that dental content is a good fit for conversational user interfaces when it is carefully curated and monitored.

Furthermore, healthcare reviews of chatbot applications have identified booking of appointments, remote patient support, and outreach to disadvantaged/equalities groups as sub-areas of benefits [1, 3]. In contrast, some inclusive-design literature has called out the need for improved support for accessibility, multimodal input, and user-specific customization options. At the same time, technical systems design work in the AI chatbot literature has provided some evidence for multimodal input, large-language-model backends, and integrated appointment scheduling modules [9, 13]. Collectively, these sources provide a basis for designing a virtual receptionist into an end-to-end workflow for dental clinics.

The purpose of the present article is to combine the conceptual framework derived from this body of literature with empirical evidence from real-world deployment of an AI-driven virtual receptionist (DentalAssist.ai) across ten dental clinics. The study pursues three objectives:

- To systematise current evidence on healthcare chatbots, AI scheduling systems and dental conversational AI, with a focus on functions relevant to reception workflows.
- To construct a conceptual architecture for an AI-driven virtual receptionist that integrates multimodal conversational capabilities with dental scheduling, triage of administrative requests and human oversight.
- To analyse, on the basis of 3,532 anonymised calls and 66 extracted turn-by-turn dialogue transcripts (50 analysable) collected over nine months of commercial operation, the workflow effects, intent distribution, after-hours coverage and conversational patterns of an AI virtual receptionist deployed in dental practice settings.

The novelty of the work lies in combining general healthcare chatbot and AI scheduling evidence with dental-specific research and with operational data from a multi-clinic deployment, producing a domain-adapted, practice-oriented framework that is conceptually grounded and empirically anchored. The article does not present a single-site case study. It uses a multi-site dataset, together with an analytical synthesis of recent literature, to map a design space and outline implementation patterns suitable for commercial dental AI products and in-house innovation in digitally mature practices.

This article advances current work on dental practice automation in five directions:

- It introduces an AI virtual receptionist framework for dental practices that integrates conversational triage with scheduling workflows and human oversight.
- It outlines, as a design specification for future work, an optimisation model for appointment allocation in multi-chair dental practices; the model is not solved or empirically evaluated in this study.
- It derives a unified workflow model for intake, triage, scheduling, and follow-up based on the empirical and conceptual literature on healthcare chatbots, AI scheduling, and dental conversational systems reviewed in Section II.

- It connects operations research concepts such as priority queues, constrained optimisation and capacity utilisation with efficiency issues in dental reception work, including chair-time wastage, reception bottlenecks and fragmented communication.
- It supplies empirical evidence from a real-world multi-site deployment, contributing one of the first practice-grounded operational profiles of a voice-based AI receptionist in dental settings.

The empirical component of this study draws on operational data collected from clinics using the DentalAssist.ai virtual receptionist as part of routine practice. The dataset captures patient interactions across multiple clinics over an extended period and provides observational evidence of how an AI receptionist operates in a live commercial context, in place of a controlled experimental evaluation.

II. MATERIALS AND METHODS

This study combines a literature-grounded conceptual framework with retrospective analysis of operational data collected during the commercial operation of an AI-driven virtual receptionist. The conceptual component adopts a narrative analytical review methodology. The empirical component follows an observational multi-site design based on call logs and anonymised dialogue transcripts.

2.1 Conceptual Framework: Literature Base

Four themes organise the reviewed literature. The first covers administrative and patient-facing functions of healthcare chatbots: scoping and rapid reviews document appointment management, information provision, health education and lifestyle support as core deployment areas, alongside economic considerations and barriers to adoption [1], [2], and a systematic review of post-intervention conversational agents reports high uptake (an estimated 97% for system-initiated questions) and sustained engagement across surgical and chronic-disease follow-up [14].

The second theme concerns scheduling optimisation. A metanarrative review of AI patient scheduling in real-world health systems catalogues machine learning and optimisation models aimed at no-show prediction, dynamic slot management and prioritisation of acute visits, with reported gains in provider productivity [4]. Classical operations research work on outpatient appointment scheduling, queueing analysis and planning taxonomies supplies the formal backbone for these systems [5–8].

The third theme is dental-specific evidence. Chatbots answer prosthodontic patient FAQs with high though uneven quality and readability [10]; comparisons of chatbot and dentist answers to dental questions find similar informativeness but weaker empathy in machine output [11]; and dental students who trained with a chatbot-based virtual patient reported satisfaction and perceived gains in diagnostic skill [12].

The fourth theme spans accessibility and system architecture. A five-year review of inclusive healthcare chatbots maps their classification, engagement and accessibility gaps [3]. Architectural papers describe deployable builds: a person-centred VHC-Bot that combines large language models, diagnostic algorithms and workflow integration under clinician supervision [9], and a multimodal AI healthcare chatbot with an integrated appointment module [13].

The review prioritised papers published in peer-reviewed sources from 2021 to 2025, with one paper from 2022 in the field of dental education retained due to its close relevance; the review also draws on classical operations research sources on appointment scheduling and queueing theory published between 2003 and 2012 [5–8], which remain foundational for the optimisation perspective. The literature analysis follows three steps: extraction of functions, design patterns, outcome measures and limitations of healthcare chatbots and AI scheduling solutions from the cited works; comparative synthesis of features relevant to dental front-desk activities, including appointment handling, information provision, triage and inclusivity; construction of a conceptual AI-driven virtual receptionist architecture and workflow model adapted to dental practice scenarios.

2.2 Empirical Component: Study Design

The empirical component is an observational, multi-site retrospective study based on operational data generated by the DentalAssist.ai virtual receptionist during commercial deployment. No experimental manipulation, randomisation or controlled comparison was performed. The analysis describes patterns of call traffic, intent and outcome that occurred over the observation window. The design aligns with the use of real-world data for healthcare AI evaluation discussed by Knight et al. [4] and reflects that a controlled trial of a commercial reception system is unfeasible at the scale considered here.

2.3 Setting and Participating Clinics

The study covers 10 dental practices in North America that adopted DentalAssist.ai to handle incoming calls during the observation period (June 2025 – March 2026). Two clinics, referred to throughout as Clinic A and Clinic B, generated 2,951 and 478 calls; the remaining sites contributed 103 calls in total. The remaining eight clinics, designated as pilot sites, contributed 103 calls in aggregate and reflect early-stage deployments before reaching full operational

use. The empirical analysis focuses on Clinics A and B, with pilot sites described at the aggregate level. The sample is neither random nor representative of dental practices in general. It reflects clinics that adopted a single commercial system, from which anonymised, aggregated operational data were analysed.

2.4 Data Sources

2.4.1 Call Log Dataset

The primary dataset comprises 3,532 anonymised call records spanning 296 days (3 June 2025 – 26 March 2026, around 9.9 months). Each record contains 29 attributes structured as shown in Table 1.

Table 1: Structure of the operational call log dataset

Attribute group	**Fields included**
Identifiers and timestamps	Call_ID, Clinic_ID, Call_Start_Timestamp, Created_At, Day_of_Week, Hour_of_Day, Call_Duration_Seconds
Caller and channel	Caller_Phone, Caller_Name, Caller_Email, Caller_DOB (all redacted), Conversation_Language, Lead_Source
Intent and dental context	Intent_Category, Intent_Subcategory, Dental_Issue, Appointment_Type, Emergency_Flag, New_Patient_Flag, Existing_Patient_Flag
Outcome and operational flags	Outcome_Status, Booking_Completed_Flag, Cancelled_Flag, Rescheduled_Flag, Escalated_To_Human_Flag (unpopulated across all 3,532 records), Callback_Requested_Flag, Call_Abandoned_Flag, After_Hours_Flag, Lead_Tag

All personally identifiable information (PII), caller name, phone number, email and date of birth, was redacted at source before analysis, consistent with privacy-preserving data handling practices commonly applied in healthcare environments. All clinic and patient identifiers were removed prior to analysis, results are reported in aggregate form only, and no individual clinic is identified in this publication. Free-text fields (Dental_Issue, Lead_Tag) were retained because they did not contain PII and were needed for intent characterisation.

2.4.2 Dialogue Transcripts

To complement the structured log analysis, a sample of 66 anonymised turn-by-turn dialogue transcripts was extracted from five export files (call-logs-2026-04-15_*.csv) covering interactions handled by the AI receptionist. Transcripts include caller utterances and AI responses verbatim, along with the system-assigned dental issue, lead tag, and source channel. All dialogues were fully de-identified and contained no patient personal data. Of these, 50 records contained complete dialogue text suitable for qualitative analysis (mean length: 844 characters; mean patient turns: 4.7, median 4). Transcripts served to characterise dialogue structure, identity verification practices, emergency handling and escalation triggers. The transcripts form a non-random convenience sample drawn from available export files and over-represent emergency-related content relative to the full dataset; the paper reports transcript-derived counts as observed patterns, not prevalence estimates.

2.4.3 Operational Dashboards

Operational dashboards exported from the management console provide clinic-level summary reporting. These were not used as an analytic data source; all figures in this paper derive from the call log dataset and the transcripts.

2.5 Pre-implementation Baseline

Formal call-handling logs from the period preceding AI implementation are unavailable for the participating clinics. The pre-implementation reference point relies on practitioner estimates gathered during routine platform use; no structured instrument, respondent count or collection date was documented. According to these estimates, before AI implementation, 80–120 calls per month went unhandled, with an estimated 40–60% follow-up gap for after-hours interactions. These figures reflect operational experience and are used in this study as a directional reference for after-hours coverage. No statistical comparison is made with post-implementation metrics.

2.6 Variables and Operational Definitions

Calls were classified along three dimensions. First dimension is intent, derived from the system-assigned Intent_Category and Intent_Subcategory fields. For analysis, intent values were consolidated into seven categories: Appointment Booking or Scheduling, Rescheduling or Cancellations, Emergency or Urgent Concerns, General Inquiries / Information Request, Human Assistance / Callback, Existing Appointment Inquiry, and Unclear or Incomplete Intent. Marginal categories with fewer than 10 calls, such as Order & Logistics, Communication & Collaboration, Materials & Products, were grouped under "Other".

The second dimension is outcome, taken from the Outcome_Status field with three values: Success (the AI completed the requested task or captured the lead),

Incomplete (the interaction terminated before resolution), and Other (residual cases, including voicemail and ambiguous terminations).

The third dimension is operational flags, boolean indicators for booking completion, cancellation, rescheduling, callback request, call abandonment, after-hours occurrence and emergency status. The After_Hours_Flag was assigned according to clinic-specific operating hours. The Emergency_Flag was set by the system from intent classification combined with red-flag keyword detection.

The Escalated_To_Human_Flag field was present in the schema but unpopulated across all 3,532 records. Human escalation was therefore measured indirectly through three proxy variables: a Lead_Tag value of "call_back", a Callback_Requested_Flag set to "true", and an Intent_Category of "Human Assistance or Call backs". The combined proxy provides a lower bound for human-handover events. This measurement limitation is addressed in the Discussion.

2.7 Analytical Methods

Descriptive analyses included frequency distributions and proportions for categorical variables, summary statistics (mean, median, standard deviation) for call duration, and cross-tabulations of intent by outcome and by after-hours status. Booking conversion rates were calculated using three denominators to remove ambiguity: all calls in the dataset, calls that reached a Success outcome, and calls with booking intent. Reporting all three values acknowledges that conversion is sensitive to scope assumptions and supports comparability with prior literature, which uses heterogeneous denominators [4]. Inter-clinic comparison was restricted to the two principal clinics - A and B. Pilot sites were aggregated. Temporal analysis examined call distributions by hour of day and day of week, with after-hours traffic isolated for separate description. Inspection of the raw exports found mixed timestamp formats, including nine records with explicit UTC markers, and a seasonal shift in the within-clinic hourly profile that daylight-saving transitions do not explain. Consistent normalisation of Call_Start_Timestamp to clinic-local time is unconfirmed, so the paper reports hour-of-day figures as recorded and attaches no behavioural interpretation to them.

Qualitative analysis of the 66 transcripts followed a content-coding approach. Each transcript was reviewed against a predefined coding frame covering: opening turn and identity verification; intent capture; booking flow steps, where applicable; emergency triage handling; and the presence of escalation triggers, i.e., caller request for a human, system-initiated callback, voicemail handover. Aggregated counts of pattern occurrence appear in the Results section. Coding was performed by a single coder. Inter-rater reliability was not assessed. This is noted as a limitation. A rule-based coding pass over the 50 transcript texts, run independently of the system's own labels, provides a partial and reproducible check on system labelling: coded emergency-related content matched the system's dental-issue summaries in 50 of 50 transcripts (Cohen's $\kappa = 1.00$), and coded human-handover triggers matched the system's call_back lead tag in 45 of 50 ($\kappa = 0.49$), with the disagreements confined to dialogues that mention a callback without carrying the tag. The check covers transcript-level fields only; human coding of a 200–300-call sample keyed to Call_ID remains future work.

2.8 Ethics, Confidentiality and Data Governance

The dataset was generated as part of the normal commercial operation of a dental call-handling service, and only anonymised operational records were made available for analysis. All identifying fields were redacted at source before analysis, and the author had no access to identifiable patient data at any stage. Free-text fields were inspected for residual PII. None was found. Because the analysis used pre-anonymised operational data that could not be linked to individual patients, it did not constitute human-subjects research requiring formal ethical review. The conduct of the study was consistent with privacy-preserving data handling practices commonly applied in healthcare environments. No clinic is named in this paper, and no clinic-level finding is attributed to an identified practice.

The architecture of the system under study is described in the Discussion at the level summarised in publicly available materials by the system operator: a hybrid AI architecture that combines large language models, real-time speech recognition, intent detection tuned to dental workflows, and rule-based orchestration for safety and operational control. Proprietary implementation aspects fall outside the scope of this article.

2.9 Limitations of the Empirical Design

Three limitations of the empirical design merit mention here, with further discussion in Section V. First, the absence of a formally logged pre-implementation baseline restricts the analysis to post-implementation description. Before-and-after comparison rests on practitioner-reported estimates. Second, the concentration of call volume in two clinics constrains the generalisability of inter-clinic patterns. The eight pilot sites are too small individually for statistical comparison. Third, the unpopulated Escalated_To_Human_Flag field forced the use of proxy measures for human handover, yielding a lower bound on escalation events rather than a precise count.

III. CONCEPTUAL FRAMEWORK

Across the reviewed settings, healthcare chatbots perform two families of functions relevant to dental reception: remote, patient-facing tasks and administrative support for professionals [1–3]. Prosthodontic FAQ evidence shows that general-purpose chatbots answer most patient questions at high quality, with Gemini scoring 4.58 out of five against 3.73 for Copilot and harder-to-read output from ChatGPT [10]. Post-intervention agents sustain high uptake and retention [14], inclusive-design reviews place appointment access among the leading chatbot use cases [3], and AI scheduling research contributes no-show prediction, dynamic slot management and priority handling of acute visits [4]. Published architectures, including a modular AI healthcare chatbot with an optimisation-based appointment module [13] and the person-centred VHC-Bot [9], supply concrete building blocks for a dental receptionist.

Drawing on these patterns, the proposed virtual receptionist framework for dental practices comprises four functional layers: multichannel intake; conversational understanding and triage; scheduling and workflow orchestration; oversight, logging, and analytics.

The multichannel intake layer supports web chat, mobile messaging, voice calls, and potentially kiosk-based input in the clinic. The conversational layer uses a domain-tuned large language model constrained by templates and retrieval from a curated knowledge base, for example, structured explanations of consent forms or post-operative care instructions. The scheduling layer connects to the practice-management system and applies rules derived from AI scheduling research, including priority queues for urgent symptoms, constraints on the sequence and spacing of complex procedures, and predictive no-show scores inspired by the models reviewed by Knight et al. [4]. The oversight layer records each interaction, provides a human-readable summary for clinicians, and supports periodic audit of dialogue snippets for quality and safety.

3.1 Optimisation Model for Appointment Allocation (Design Specification for Future Work)

This subsection presents a design specification. The study does not parameterise the model from the dataset, solve it, or evaluate it against the observed request stream; that work remains for a follow-up study. The scheduling module is formulated as a finite-horizon appointment allocation problem. Let I be the set of patient requests and T the set of available appointment slots. For each patient $i \in I$ and slot $t \in T$, the binary variable $x_{i,t}$ takes value 1 if patient i is assigned to slot t and 0 otherwise. The parameter d_i denotes the predicted appointment duration for patient i obtained from the triage module, and p_i denotes a priority score generated from symptom severity, waiting time, and contractual constraints. For each slot t, s_t indicates whether the slot is open for booking, while c_t describes the chair and provider capacity available during that slot.

The optimisation objective is to allocate requests to slots so that unused capacity and schedule fragmentation are reduced while throughput and service levels improve. A representative multi-criteria objective can be written as:

$$min \ \sum_t \alpha \cdot idle_t + \sum_t \beta \cdot gap_t - \sum_{i,t} \gamma \cdot p_i \cdot x_{i,t} + \sum_i \delta \cdot delay_i \quad (1)$$

where $idle_t$ measures predicted idle chair time in slot t, gap_t quantifies undesirable gaps between appointments, and $delay_i$ captures waiting time until the allocated appointment for patient i. The weights α, β, γ, and δ are set by the clinic to balance utilisation, access, and equity.

The decision variables are subject to operational constraints reflecting dental practice realities. Each patient is assigned to at most one slot:

$$\sum_{t \in T} x_{i,t} \leq 1 \qquad \forall i \in I \quad (2)$$

and the total occupied time in a slot respects chair and provider capacity:

$$\sum_{i \in I} d_i \cdot x_{i,t} \leq c_t \qquad \forall t \in T \quad (3)$$

Slots flagged as unavailable are excluded by enforcing:

$$x_{i,t} = 0 \qquad whenever \ s_t = 0 \quad (4)$$

Additional constraints express emergency handling, buffer requirements and business rules. Emergency requests with high p_i values receive access to reserved emergency slots and may trigger reallocation of elective appointments. Minimum buffer times between procedures are enforced by scheduling incompatible appointments at different times. Fixed opening hours, breaks and provider-specific templates are represented by restricting T and c_t.

The interaction between AI and optimisation is explicit. The conversational and triage modules estimate d_i, p_i, and no-show probabilities, classify request types, e.g., hygiene visit, restorative treatment, new patient consultation, and collect patient preferences. These quantities feed the optimisation engine, which runs either in real time during the dialogue or in periodic batches. The constraint evaluation module checks the feasibility of candidate schedules given clinic capacity and policy rules, while the optimisation engine searches for the schedule that minimises the objective. The selected appointment proposal is then returned through the appointment slot selector in the AI decision layer, presented to the user and written back into the practice-management system once confirmed.

The conceptual architecture follows the data-flow patterns of the AI Healthcare Chatbot [13], specialised to dental scheduling: the system transcribes text or voice input,

a domain-constrained language model classifies intent, scheduling intents query the practice-management database under configurable chair and workload rules, and red-flag phrases route the conversation to staff. Dental evidence supports this design: chatbots reach acceptable informational quality under oversight [10], structured conversational flows aid diagnostic training [12], and inclusive-healthcare reviews direct attention to modalities, languages and accessibility [3].

IV. RESULTS

This section reports empirical findings from the operational dataset, organised in five parts: (i) overall traffic profile, (ii) temporal distribution, (iii) intent distribution and outcomes, (iv) inter-clinic comparison, and (v) qualitative analysis of dialogue transcripts. All figures are derived from the 3,532-call dataset and the transcripts described in Section II (66 extracted, 50 analysable).

4.1 Overall Traffic Profile

Table 2 summarises the volume and outcome metrics for the full 9.9-month observation window. The system handled 3,532 calls across 10 dental practices, with 67.6% reaching a Success outcome and 28.1% terminating without resolution. One in seven calls (15.1%) was abandoned, and 27.3% occurred outside standard business hours.

Table 2: Overall traffic profile of the AI virtual receptionist over the observation period

Metric	Value	% of total
Total calls	3,532	100.0%
Observation period	296 days	≈ 9.9 months
Active clinics	10 sites	2 primary, 8 pilot
After-hours calls	964	27.3%
Abandoned calls	533	15.1%
Calls reaching Success outcome	2,389	67.6%
Calls with completed booking	743	21.0%
Emergency-flagged calls	185	5.2%
Mean call duration (s)	95.9	median 39.5 s
Calls in English (of identified-language)	2,705 / 2,725	99.3%

Two structural features of the dataset shape the interpretation of the metrics that follow. First, English dominates the language profile. Among the 2,725 calls with an identified-language tag, 99.3% were conducted in English, with 17 calls in Hindi, 2 in Korean and 1 in Italian. The system's multilingual capability is therefore not stress-tested in the current sample. Second, the mean call duration was 95.9 seconds (median 39.5), indicating that successful interactions are short and task-focused. Long-tail durations (maximum 2,442 seconds) likely correspond to complex emergency or rescheduling dialogues.

4.2 Temporal Distribution

Call traffic concentrates on weekdays. As shown in Fig. 1, Friday accounted for 29.5% of weekly calls (n = 1,042), followed by Wednesday at 23.6% (n = 832). Together, these two days carried 53.1% of the total weekly volume. Weekend traffic was negligible: Saturday and Sunday combined contributed 0.8% of all calls (n = 30). The pattern aligns with weekend closures across dental practices and indicates that the AI receptionist absorbs only a small share of out-of-business-day demand in this sample.

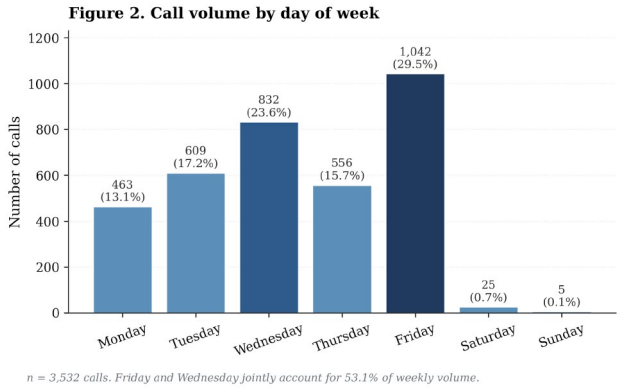


Fig. 1: Call volume by day of week

Within the weekday cycle, call volume peaks in the late morning. The 10:00–12:00 window accounted for 31.1% of all calls (n = 1,099), with sub-peaks at 10:00 (399), 11:00 (350), and 12:00 (350). Volume tapered through the afternoon and rose again in the early evening (18:00–19:00: 238 calls, 6.7%). Traffic recorded at 06:00–07:00 totalled 471 calls (13.3%). Fig. 2 displays the full hourly distribution and identifies two after-hours clusters: an early-morning cluster (04:00–07:00) and a smaller early-evening cluster (18:00–19:00).

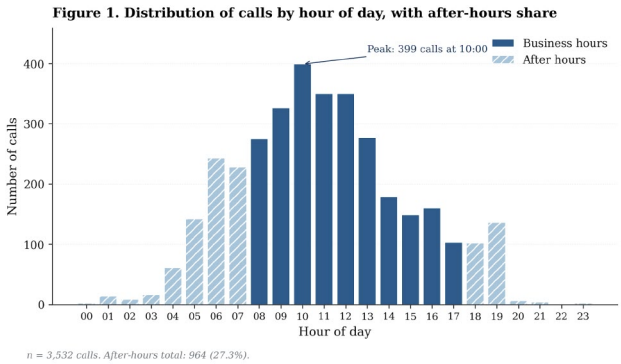


Fig. 2: Distribution of calls by hour of day, separating business-hours (solid) and after-hours (hatched) traffic

4.3 After-hours Coverage

Of the 3,532 calls, 964 (27.3%) occurred outside the participating clinics' standard operating hours. Within this after-hours subset, 645 calls (66.9%, 95% CI 63.9–69.8%) reached a Success outcome, 261 (27.1%) were Incomplete, and 238 (24.7%, 95% CI 22.1–27.5%) ended with a confirmed booking; the match with the 238 calls logged in the 18:00–19:00 window is a coincidence. The success rate for after-hours calls is close to that observed during business hours (66.9% versus 67.9%). The two windows differ in clinic composition, so this pooled comparison is not adjusted for site and should be read as descriptive rather than as evidence of equivalent resolution quality across time of day. The majority of calls were received during the early morning (04:00-07:00, 674 calls, 69.9%) and early evening (18:00-19:00, 238 calls, 24.7%).

Stratifying by clinic removes the composition effect. Within Clinic A, after-hours calls reached Success in 59.7% of cases (433 of 725; 95% CI 56.1–63.2%) against 64.6% during business hours (1,438 of 2,226; 95% CI 62.6–66.6%), a deficit of 4.9 percentage points ($\chi^2(1) = 5.6$, $p = 0.018$, $\varphi = 0.04$). Within Clinic B the two windows do not differ: 92.6% (187 of 202) against 93.8% (259 of 276). Clinic B supplies 21% of after-hours calls but 12% of business-hours calls, so the pooled figures weight the higher-performing clinic more heavily outside opening hours and overstate after-hours performance. The corrected reading is that after-hours handling is close to business-hours handling within each clinic, with a small but measurable deficit at the higher-volume site.

According to the baseline reported by practitioners prior to implementation, 80 to 120 calls per month were not met, and 40 to 60% of cases were missed after-hours (Section 2.5). The post-implementation observation of 964 after-hours calls received and processed without staff involvement over 9.9 months, averaging 97 per month, falls within the same order of magnitude as the pre-implementation gap reported by clinic operators. The baseline is not formally logged, so no statistical test is performed. The comparison serves as a directional indicator.

4.4 Intent Distribution

Fig. 3 shows the consolidated distribution of intent categories across the dataset. Appointment Booking or Scheduling is the dominant category at 51.8% of all calls (n = 1,830). The second-largest category, at 29.9% (n = 1,056), comprises calls classified as Unclear or Incomplete Intent, in which the system could not assign a confident intent label, often due to early call termination, background noise, or absent caller speech. The remaining categories, General Inquiries, Human Assistance / Callback, Emergency, Rescheduling, and others, together accounted for 18.3% of traffic.

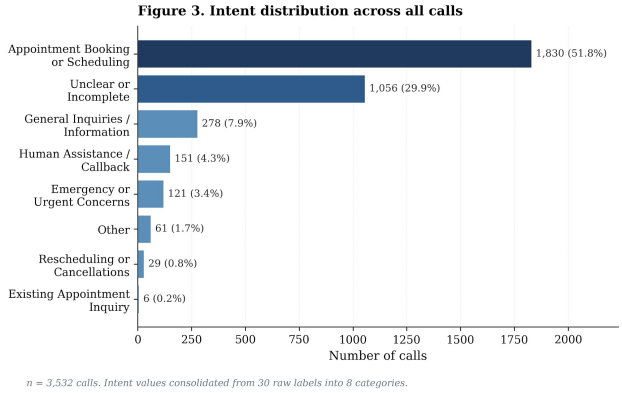


Fig. 3: Intent distribution across all calls

The Unclear or Incomplete Intent category overlaps with the abandonment category. Of the 1,056 unclear-intent calls, 502 (47.5%) were also flagged as abandoned, accounting for 94.2% of all 533 abandoned calls in the dataset. The median duration of abandoned calls was 18 seconds (mean 36.7 s), well below the dataset median of 39.5 seconds, indicating that these are short interactions that end before the dialogue progresses to intent capture. A share of "unclear" cases reflects callers hanging up, and the available fields are unable to distinguish this from genuine dialogue failures.

Emergency-flagged calls (n = 185 by Emergency_Flag, 5.2% of the dataset; n = 121 by consolidated intent category, 3.4%) showed a high success rate: 98.9% of emergency-flagged calls reached a Success outcome with no abandonments and no Incomplete terminations. Only 17.3% of emergency-flagged calls (n = 32) ended with a system-completed booking. The emergency workflow captures patient details for a clinic call-back instead of autonomous slot assignment. The mean duration of emergency interactions was 142.3 seconds, 48% longer than the dataset average, reflecting the additional information captured, such as caller name, phone number, and urgency confirmation in these dialogues. If the system sets Emergency_Flag only after intent classification and detail capture succeed, the near-perfect profile of this subgroup reflects selection by construction as much as end-to-end performance.

4.5 Booking Conversion

Booking conversion appears under three denominators (Fig. 4) for comparability with prior literature, which uses heterogeneous scope assumptions. Each denominator group carries its own booking count, so the numerator varies with

scope. Across the broadest denominator, all 3,532 calls in the dataset, 743 completed bookings give 21.0% (95% CI 19.7–22.4%). Among the 2,389 calls that reached a Success outcome, 718 completed bookings give 30.1% (95% CI 28.2–31.9%). Among the 1,830 calls with a booking intent, the most operationally meaningful denominator, 707 completed bookings give 38.6% (95% CI 36.4–40.9%). The full intent × outcome × booking cross-tabulation is provided in Table 7.

Table 7: Consolidated intent × outcome cross-tabulation with system-completed bookings (full dataset)

Consolidated intent	n	Success	Incompl.	Other	Bookings
Booking / Scheduling	1,830	1,799	3	28	707
Unclear / Incomplete	1,056	42	960	54	1
General inquiries	221	185	7	29	4
Human assistance	151	139	2	10	1
Emergency	121	120	0	1	5
Other	153	104	19	30	25
Total	3,532	2,389	991	152	743

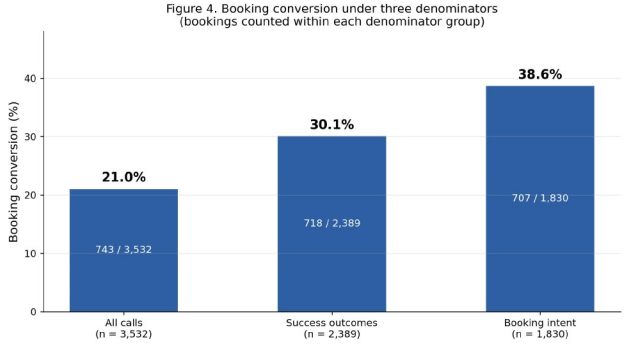


Fig. 4: Booking conversion under three denominators

When a caller arrives with the intention to book, the AI receptionist completes the booking autonomously in just under 2 of 5 cases in the observed sample. The remaining booking-intent calls did not result in a same-call booking for one of three reasons: the caller was offered slots but did not confirm before the dialogue ended, the caller requested a callback for human assistance, or the system captured the lead and deferred booking to clinic staff.

4.6 Inter-clinic Comparison

As Section 2.3 noted, call volume is concentrated in two clinics, with the eight pilot sites contributing a small residual. Fig. 5 contrasts the operational performance metrics of the two principal clinics and the aggregated pilot sites. Table 3 reports volume and structural indicators not displayed in Fig. 5.

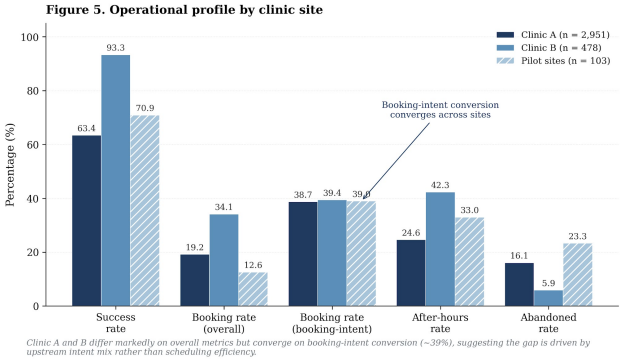


Fig. 5: Operational profile by clinic site

Table 3: Volume and structural indicators by clinic site

Metric	Clinic A	Clinic B	Pilot sites
Total calls	2,951	478	103
Incomplete rate	31.7%	5.9%	28.2%
Mean duration (s)	108.1	26.0	63.4

The two principal clinics show contrasting profiles. Clinic A handles more volume (2,951 calls) at lower success (63.4%, 95% CI 61.6–65.1%) and booking (19.2%) rates, and with a longer mean duration (108.1 seconds). Clinic B handles fewer calls (478) at a higher success rate (93.3%, 95% CI 90.7–95.2%), a higher booking rate (34.1%) and a shorter mean duration (26.0 seconds, median 0 seconds). The shorter median for Clinic B, combined with its 42.3% after-hours share, suggests that a large portion of Clinic B traffic consists of brief automated interactions resolved without extended dialogue, a pattern that fits a deployment configured for after-hours call capture. The difference in Success rates between Clinics A and B is statistically significant ($\chi^2(1) = 167.9$, $p < 0.001$, $\varphi = 0.22$). This unadjusted difference is, for the greater part, a calendar-time effect. Clinic B logged 10 calls in June–August 2025 and became active in September, whereas Clinic A logged 915 calls in those three months at a Success rate of 23.5% (215 of 915); Clinic A's aggregate therefore carries the whole pre-maturation period described in Section 4.8 and Clinic B's does not. Restricting both clinics to their common window, September 2025 to March 2026, narrows the Success gap from 29.9 to 11.8 percentage points (Clinic A 81.3%, 1,656 of 2,036, 95% CI 79.6–83.0%; Clinic B 93.2%, 436 of 468, 95% CI 90.5–95.1%; $\chi^2(1) = 38.7$, $p < 0.001$, $\varphi = 0.12$), and booking-intent conversion in the same window is identical at the two sites (40.3% and 40.4%). A

residual site difference in Success remains after this adjustment; the headline gap, however, is mostly a period effect. Of Clinic B's 478 calls, 257 (53.8%) carry a zero recorded duration at the telephony layer. Success is 87.9% among the zero-duration calls and 99.5% among the remaining 221 calls, so zero-duration records alone do not produce the clinic's aggregate 93.3%. A technical definition of zero-duration calls at the telephony layer is still missing.

When conversion is normalised to booking-intent calls only, the two clinics converge: 38.7% for Clinic A versus 39.4% for Clinic B. This convergence indicates that the gap in the overall booking rate derives from the upstream intent mix — which calls reach the system — and that booking-completion efficiency, once intent is captured, is comparable across sites. Pilot sites show similar booking-intent conversion (39.0%) but higher abandonment (23.3%), indicating that the system handles booking intent consistently across deployments and remains vulnerable to early termination in low-volume settings.

Section 4.8 assesses temporal stability directly from the dataset.

4.7 Qualitative Analysis of Dialogue Transcripts

Of the 66 transcripts extracted, 50 contained complete dialogue text suitable for content analysis. The other 16 records were short fragments, system-state messages or duplicate exports. Mean transcript length was 844 characters, with a mean of 4.7 patient turns per dialogue (median 4). Lead tagging of the 50 valid transcripts concentrated in two categories: Positive_Lead (44 transcripts, 88%) and call_back (5 transcripts, 10%), with one untagged record.

Content coding of the 50 valid transcripts identified five recurring patterns, summarised in Table 4.

Table 4: Frequency of dialogue patterns in 50 valid transcripts

Dialogue pattern	**Transcripts (n = 50)**	**Share**
Slot offering present (any time/date proposal)	13	26%
Emergency-related dialogue content	30	60%
Identity verification step (DOB / full name)	14	28%
Explicit booking confirmation utterance	9	18%
Human-handover / callback trigger	7	14%

Three findings from the qualitative analysis merit emphasis. First, the booking flow follows a four-stage structure across booking-intent dialogues: opening greeting from the AI receptionist named "Reeta AI" in the deployed configuration; classification of caller status, new versus existing patient; identity verification, full name and date of birth for existing patients; slot offering, selection and confirmation. The full sequence appears in 9 of the 50 transcripts. A time- or date-specific slot proposal appears in 13 of the 50 transcripts; the remaining dialogues ended at the intake or triage stage before any slot proposal, or involved no booking request.

Second, emergency-related interactions follow a distinct pattern that departs from the standard booking flow. In emergency cases, bleeding, severe pain or system-flagged "dental emergency", the AI receptionist confirms urgency, captures caller's name and phone number, and signals that clinic staff will contact the caller. Autonomous booking does not occur in these cases. The dialogue prioritises rapid information capture over slot assignment. The pattern aligns with the quantitative finding that only 17.3% of emergency-flagged calls result in a system-completed booking.

Third, human-escalation triggers appear in 7 of the 50 analysed transcripts, concentrated in three situations: a caller requesting a representative or callback; cases involving complaints or sensitive enquiries that the system is not authorised to resolve; handover via voicemail when the dialogue cannot complete within the AI's operational scope. These observed triggers align with the system's stated escalation policy: routing to human staff when the request is complex or sensitive, when classification confidence is low, when the caller requests a human, or when the workflow requires manual intervention.

Taken together, the qualitative findings indicate that the AI receptionist operates within a defined and observable conversational scope: a structured booking flow for routine appointments, a separate triage-and-capture protocol for emergencies, and handover paths for cases outside its remit. The patterns are consistent across the two principal clinics within the available sample.

4.8 Temporal Stability and Learning Effects

Table 8 reports monthly Success, booking-completion and abandonment rates across the full 296-day window. The series traces a steep maturation curve: in the first three months of the window (June–August 2025, n = 978) the Success rate was 26.2%, booking completion 3.3% and abandonment 21.0%, whereas in the final three months (January–March 2026, n = 1,308) Success reached 86.2%, booking completion 32.8% and abandonment fell to 9.4%. The observation window coincides with active product development, so month-level figures confound patient

behaviour with configuration and model changes; the aggregate values reported elsewhere in this section (67.6% Success, 21.0% booking completion) average over this trajectory and understate the performance of the mature system while overstating that of the early deployment. The series is not monotonic at the end of the window: March 2026, the highest-volume month in the dataset, records a lower Success rate (77.8%) and booking-completion rate (25.7%) than either January or February. The cause is not established from the available data and the decline is reported as unexplained variance.

Table 8: Monthly operational metrics across the observation window (all clinics)

Month	Calls	Success	Booking	Abandoned
2025-06	94	51.1%	2.1%	30.9%
2025-07	673	17.4%	4.0%	12.3%
2025-08	211	43.1%	1.4%	44.1%
2025-09	353	83.0%	19.0%	16.7%
2025-10	438	78.5%	18.3%	17.1%
2025-11	182	80.8%	39.0%	13.2%
2025-12	273	81.0%	23.4%	17.2%
2026-01	370	91.4%	41.9%	8.4%
2026-02	420	92.1%	33.6%	6.2%
2026-03	518	77.8%	25.7%	12.7%

V. DISCUSSION

This study combines the conceptual framework derived from healthcare chatbot, AI scheduling and dental conversational AI literature with empirical evidence from a multi-clinic deployment of an AI virtual receptionist. The discussion is organised in four parts: interpretation of the empirical findings against the published literature; implications for the proposed framework and architectural choices; practical implications for dental practices; and limitations and directions for future work.

5.1 Empirical Findings in the Context of the Published Literature

The synthesised framework positions the virtual receptionist as a central orchestrator for communication and scheduling in dental practices. The empirical patterns observed in this study support that positioning and supply practice-grounded magnitudes of a kind so far absent from the dental AI literature.

The dominance of Appointment Booking or Scheduling intent (51.8% of all calls) and the share of Emergency or Urgent Concerns (3.4% by consolidated category, 5.2% by Emergency_Flag) confirm that scheduling and triage are the two principal workloads for a dental virtual receptionist, in line with the function inventories reported by M. Barreda et al. [1] and M. Laymouna et al. [2]. The 38.6% booking conversion rate for booking-intent calls is, to the author's knowledge, one of the first published quantitative benchmarks for AI-assisted booking in real-world dental practice. Direct comparison is constrained by heterogeneous denominators in the prior literature. The magnitude aligns with the metanarrative review of AI scheduling systems by D.R.T. Knight et al. [4], which reports productivity gains and reduced provider burden across implementations using comparable real-world designs.

Healthcare chatbot reviews report that administrative and access-related uses are common. The scoping review by M. Barreda et al. [1] identifies appointment management as one of several recurring chatbot functions alongside mental health support, information delivery and lifestyle change support. In a rapid review, M. Laymouna et al. [2] divide chatbots into two categories: remote service delivery, and administrative responsibilities like appointment scheduling, referrals and data collection. In a scoping review including non-healthcare chatbots, E. Grassini et al. [3] find that about 19% of deployed chatbots are for accessing healthcare services, like booking an appointment and accessing contact information. This suggests that the booking-intent share of 51.8% observed here is considerably higher than the 19% access-related share reported by Grassini et al. The difference is expected. The present sample is drawn from a system deployed specifically to replace a dental front-desk function, so booking-related contacts are over-represented by design relative to the general-purpose healthcare chatbots surveyed in that review.

After-hours coverage was one of the most powerful signals in the analysis. After-hours calls, 27.3% of the total, reached almost the same Success rate as business-hours calls (66.9% versus 67.9%); the within-clinic comparison in Section 4.3 shows the same closeness at both principal sites, with a 4.9-point after-hours deficit at Clinic A. This aligns with the access barriers recognised in the inclusive-healthcare review literature [3] and an access-oriented framing of multimodal chatbots [13]. The pre-implementation practitioner-reported gap of 80–120 unhandled calls per month, with 40–60% follow-up failure for after-hours interactions, falls in the same order of magnitude as the post-implementation handling of around 97 after-hours calls per month observed in this study. The comparison is directional and not statistical. It indicates that AI-driven receptionists can close the most common operational gap reported by clinic operators, overnight and

early-morning calls that historically went to voicemail or were dropped.

The quality and safety of patient-facing materials must be governed. H. Esmailpour et al. have shown that AI chatbots can produce responses of high quality to most FAQs in prosthodontics, but produce low quality responses to a small subset of these questions [10]. The same study reports that some chatbot outputs demand a high reading level and remain hard for patients to follow [10]. E.E. Maruska et al. [11] report comparable informativeness between chatbot and dentist answers, with machine output at times matching or exceeding dentist accuracy while offering less empathy and context. The empirical finding that emergency-flagged calls reached a Success outcome in 98.9% of cases while only 17.3% ended with autonomous booking matches a system constrained to capture-and-handover behaviour for sensitive cases, a design pattern that aligns with the cautionary recommendations in the dental conversational AI literature. The qualitative observation that 7 of the 50 analysed dialogues exhibited human-escalation triggers reinforces this interpretation: the system is designed to recognise its operational scope and route out-of-scope cases to human staff.

The 29.9% share of Unclear or Incomplete Intent, of which 47.5% are also abandoned, is the largest single source of unrealised potential in the dataset. The pattern of short duration, abandonment without dialogue and no confident intent assignment points to a class of interactions where intent capture itself is the bottleneck. Improving the opening turn of the dialogue (lower latency, refined prompt phrasing, graceful handling of early caller hang-ups) may yield disproportionate gains, in line with the engineering attention to prompt and dialogue design highlighted by Shaikh et al. [13] and Alsalamah et al. [9].

The scheduling aspect was discussed in the context of appointment scheduling systems [5, 6] in the metanarrative study by D.R.T. Knight et al. [4] on AI for patient scheduling systems with no-shows, dynamic sequencing of patients, and dynamic capacity planning of clinical appointments. The authors found heterogeneity in models and outcomes. AI scheduling can reduce the workload of providers and increase productivity when implemented in clinical workflows. Most reviewed systems operate in hospitals or outpatient settings instead of dental practices. The underlying principles, probabilistic modelling of attendance, optimisation of slot allocation under resource constraints, and integration with electronic systems, transfer to multi-chair dental clinics. The proposed framework incorporates predictive and optimisation logic in its scheduling layer, while leaving clinical approval and final decision-making to humans. The empirical convergence of booking-intent conversion across the two principal clinics (38.7% and 39.4%) indicates that the scheduling component performs consistently when intent is well captured, and that residual variation between sites stems from upstream factors, such as intent mix, call duration profile, and deployment configuration.

Comparing different strands of evidence clarifies how benefits cluster and where uncertainties remain. Table 5 indicates that benefits cluster around efficiency, access and experience, with direct implications for dental reception workflows.

Table 5: Reported benefits of chatbots and AI scheduling relevant to a dental virtual receptionist [1, 2, 9, 10, 12–14]

Primary setting	**Function relevant to reception**	**Reported benefit type**
Virtual hospital / virtual care	Person-centred conversational triage and coordination	Shorter consultation times, improved clinician–patient communication, higher patient satisfaction, enhanced workflow indicators
Multiple healthcare domains	Appointment management, patient information, and monitoring	More efficient delivery of remote services, potential cost savings, and improved access to services
Dental prosthodontics	Automated answering of patient FAQs	High average response quality, improved accessibility of information, and better readability for some models
Post-operative follow-up	Automated follow-up and symptom monitoring	High response rates, sustained engagement, and reduced need for in-person follow-up contact
eHealth applications	Access to healthcare, appointment booking, and remote assistance	Expanded access for remote and underserved users, potential reduction in organisational burden
Real-world healthcare scheduling	AI-based patient scheduling	Reduced provider burden, increased productivity, and opportunities for better slot utilisation
Broad healthcare landscape	Administrative assistance, remote services	Greater accessibility, potential for more efficient care pathways, and a wide range of user groups served

General healthcare	Multimodal conversational support and appointment module	Streamlined intake and appointment scheduling, reduced administrative workload, improved patient access
Dental education	Structured diagnostic and communication dialogues	High user satisfaction, perceived improvement in diagnostic reasoning and communication skills

Evidence from VHC-Bot and the AI Healthcare Chatbot indicates that conversational systems shorten consultation time while sustaining or improving patient satisfaction [9, 13]. Remote-service reviews report that chatbots sustain engagement in follow-up scenarios [1, 2, 14], and dental-specific work demonstrates that chatbots convey clinically relevant information at acceptable quality [10–12]. Set against these gains are documented limitations and risks reported across the literature (Table 6).

Table 6: Principal risks and limitations of healthcare chatbots aligned with dental virtual receptionist use [2–4, 9–11]

Risk or limitation type	**Relevance to the dental virtual receptionist**
Occasional poor-quality, misleading responses; variable readability	Requires content governance, constrained domains, and readability tuning for dental FAQs
Limited attention to accessibility and inclusion in many implementations	Necessitates explicit accessibility requirements (screen readers, plain language, multilingual support) for reception interfaces
Heterogeneous models and metrics; limited real-world evidence on long-term sustainability	AI scheduling components should be monitored for unintended effects on access and fairness; human oversight is needed for policy decisions
Ethical, medicolegal, and user-experience concerns; risk of overreliance	Dental clinics must communicate chatbot limits, log interactions, and design clear escalation to human staff
Lower empathy and inconsistent nuance compared with dentists	Virtual receptionists should be restricted to administrative and low-risk informational tasks; sensitive discussions should stay with clinicians
Risk of clinician deskilling and unsafe self-diagnosis if AI is overused	System design should preserve clinician authority and avoid autonomous clinical decision-making in scheduling and triage

5.2 Implications for the Framework and Architecture

The empirical findings reinforce three architectural choices outlined in Section III.

First, the four-layer decomposition of the framework, multichannel intake, conversational understanding and triage, scheduling and workflow orchestration, and oversight, logging and analytics, maps onto the empirical patterns in a usable way. The intake and conversational layers are stress-tested by the share of Unclear / Incomplete intent and abandoned short calls. The scheduling layer converts ≈ 39% of booking-intent calls into confirmed bookings across deployments. The oversight layer is operationalised by the structured outcome and intent fields in the dataset. The analytical telemetry envisaged by the framework is therefore realisable in production.

Second, the empirical Emergency-handling pattern, high Success rate, combined with low autonomous booking rate and elevated dialogue duration, supports the framework's emphasis on defined escalation rules and on human oversight for high-priority cases. The pattern aligns with the operational stance described by the system operator: calls escalate when the request is complex or sensitive (such as emergencies or complaints), when classification confidence is low or ambiguity is detected, when the caller requests a human, or when the workflow requires manual intervention. Emergency-related scenarios follow strict escalation protocols for patient safety. The incidence of human-handover triggers in 7 of the 50 analysed transcripts indicates that escalation is an observable component of system behaviour, not a theoretical fallback.

Third, the system architecture as described by the operator, a hybrid configuration combining large language models, accessed via providers such as OpenAI, with real-time speech-to-text processing, dental-workflow-tuned intent detection, context-aware response generation, workflow automation and rule-based orchestration for safety and operational control, matches the modular pattern proposed in the framework and the multimodal modular architectures described by Shaikh et al. [13] and Alsalamah et al. [9]. The combination of generative and rule-based

components carries weight in the empirical pattern: emergency dialogues follow a tightly bounded capture-and-handover pattern that differs from the more flexible booking flow, which indicates that rule-based orchestration enforces the safety boundaries the literature has called for.

5.3 Practical Implications for Dental Practices

From a systems-engineering perspective, dental clinics form a favourable environment for such a framework. They operate with fewer integration layers than hospitals because practice management systems oversee scheduling, billing, and clinical documentation. The AI scheduling insights reported by Knight et al. [4] can be embedded as decision-support services. A no-show prediction model, for example, proposes an overbooking range. The receptionist assistant suggests rescheduling low-priority visits without mandating it, preserving clinician and manager oversight while drawing on patterns extracted from historical data.

In this setting, the AI receptionist functions as a dispatcher for several patient queues, where requests are classified, assigned priority weights and routed to constrained resources such as chairs, providers and communication channels. Queue-based scheduling and optimisation-driven appointment allocation form a structured approach to relieving workflow bottlenecks caused by ad hoc booking rules and manual triage at the front desk. The empirical observation that Friday and the 10:00–12:00 window jointly carry a disproportionate share of weekly volume (29.5% of weekly traffic concentrated in a single weekday; 31.1% of daily traffic concentrated in a three-hour morning window) supplies an operational argument for AI-driven load smoothing and priority queuing during peak intervals.

Another dimension concerns staff experience. Front-desk teams in dental practices handle multiple concurrent tasks, including phone communication, arrivals, insurance checks and coordination with clinicians. Research on chatbots and virtual care, such as work by Alsalamah et al. [9], Barreda et al. [1], Laymouna et al. [2] and Shaikh et al. [13], shows that the automated management of routine, low-judgment tasks using a configured virtual receptionist that can process out-of-hours requests, reschedule appointments using a self-service option, and provide standard information messages reduces demand peaks. This allows staff to focus on more difficult situations, such as distressed patients or complicated insurance questions. Research on automating administrative responsibilities in healthcare has reached similar conclusions that scheduling and inquiries are good candidates for automated support [2, 4]. The empirical 27.3% after-hours share confirms that this displacement effect is observed in practice. It represents one in four interactions absorbed without consuming staff time during business hours.

Converting informal booking heuristics into transparent optimisation rules gives clinics a tunable mechanism for balancing high-priority and routine visits. Priority-weighted triage ordering moves clinically urgent or time-sensitive cases through the reception queue without permanently displacing preventive and elective care. The empirical pattern that emergency-flagged calls are uniformly resolved (98.9% Success) and routed via human-handover (only 17.3% autonomous booking) shows this priority logic operating in practice.

For dental patients in underserved, rural or mobility-restricted groups, the framework promises smoother access to appointments and information. Reviews focused on access to care and inclusivity describe chatbots' capacity to mitigate barriers when physical attendance or traditional communication channels are limited [1–3]. Adapted for dentistry, a virtual receptionist that supports text, voice, and multiple languages lowers entry barriers for individuals who experience difficulties with phone communication during working hours or find conventional booking methods inconvenient. The current dataset shows concentration in English (99.3% of identified-language calls) and a phone-first channel mix. The multilingual and multichannel capacity of the framework, remains untested at scale in this sample and is a target for future evaluation.

5.4 Limitations

Several limitations of the empirical component should be borne in mind when interpreting the findings.

First, the absence of a formally logged pre-implementation baseline restricts the analysis to post-implementation description. The practitioner-reported estimates of 80–120 unhandled calls per month and a 40–60% follow-up gap before AI deployment serve as a directional reference and cannot support a statistical before-and-after test. Future work in clinics that retain detailed historical call-handling logs would enable more rigorous comparisons.

Second, the concentration of call volume in two clinics limits the generalisability of inter-clinic patterns. The within-clinic patterns observed in Clinics A and B carry weight given the sample size. The eight pilot sites are too small individually to support per-site analysis. Aggregating pilot sites mitigates this constraint without removing it.

Third, the study does not include a controlled comparison group. The system under evaluation is a commercially deployed product, and the practical and ethical constraints of running a randomised trial of front-desk automation in active clinics are considerable.

Observational evidence from real-world deployment is the best available alternative and cannot rule out confounding factors such as concurrent practice management changes, staffing levels, or seasonal variation.

Fourth, the unpopulated Escalated_To_Human_Flag field required proxy measurement of human-handover events using a combination of Lead_Tag, Callback_Requested_Flag, and intent category. The approach yields a lower bound on escalations and does not capture cases in which staff intervened without these flags being set. Tighter telemetry on escalation events is a target for future iterations of the data layer.

Fifth, the qualitative coding of 66 transcripts was performed by a single coder against a predefined frame. Inter-rater reliability was not assessed. The aggregated counts in Table 4 should be interpreted as descriptive measurements.

Sixth, the analysis is restricted to calls handled by one commercial system within a specific deployment context. Findings on intent distribution, conversion rates, and dialogue patterns may not generalise to dental practices using different conversational AI systems, practice-management integrations, or patient populations. The framework set out in Section III is system-agnostic. The empirical evidence presented here is system-specific. Seventh, the system under evaluation assigns every analytic label in this study (Intent_Category, Outcome_Status, Emergency_Flag, Booking_Completed_Flag); no independent human coder or practice-management-system record has validated them. The findings describe system-recorded outcomes, and Booking_Completed_Flag captures system-side completion only: no check confirmed that each booked appointment existed in the practice management system and was kept. A rule-based cross-check of transcript-level fields (Section 2.7) found full agreement on emergency content and moderate agreement on callback tagging, but Intent_Category and Outcome_Status remain unvalidated. Eighth, hour-of-day figures assume that Call_Start_Timestamp and the derived Hour_of_Day follow each clinic's local time zone. Inspection of the raw exports found mixed timestamp formats, including records with explicit UTC markers, and a seasonal shift in the within-clinic hourly profile that daylight-saving transitions do not explain. The paper reports the 04:00–07:00 after-hours concentration as recorded clock time and reads no early-morning caller behaviour into it. The calls themselves are substantive: Success reaches 76.9% and booking completion 27.3% in this window, against 65.5% and 19.6% elsewhere.

Ninth, certain fields in the dataset (New_Patient_Flag, Existing_Patient_Flag) were left blank, preventing direct quantification of the new-versus-existing patient mix relevant to booking behaviour. The 28% incidence of identity-verification steps in transcripts gives indirect evidence on this dimension. A structured analysis would require populated patient-status fields.

The discussion of evidence indicates that an AI-driven virtual receptionist configured according to the proposed framework aligns with current trajectories in healthcare chatbots, schedule optimisation and digital dentistry, and that the empirical patterns observed in commercial operation align with the published literature in both magnitude and qualitative shape. The approach aligns with the direction of emerging commercial systems focused on dental communication and scheduling automation, grounded in published evaluations and now supported by real-world operational data.

VI. CONCLUSION

The article develops and empirically substantiates an AI-driven virtual receptionist framework that adapts healthcare chatbot and AI scheduling research to the operational realities of dental practices. The framework integrates multichannel conversational intake, domain-constrained large language models and a scheduling layer informed by AI-based patient scheduling studies, with provisions for human oversight, accessibility and content quality control.

The first objective, systematising current evidence, was met by combining healthcare-wide chatbot reviews, AI scheduling syntheses and dental-specific evaluations of conversational systems. The second objective, constructing a conceptual architecture, draws on recent designs for multimodal chatbot systems and person-centred virtual care platforms specialised for dental scheduling and administrative triage. The third objective, empirical evidence on workflow effects, rests on a 9.9-month observational dataset of 3,532 calls and 66 extracted dialogue transcripts (50 analysed) across ten dental sites, two of which were studied in depth.

The empirical findings indicate that in commercial deployment the AI virtual receptionist absorbs a measurable share of administrative call volume (67.6% Success outcomes overall, 38.6% booking conversion among the 1,830 booking-intent calls, 27.3% after-hours coverage), keeps a separation between routine booking workflows and emergency triage (the latter resolves in 98.9% of cases through capture-and-handover instead of autonomous booking), and exhibits identifiable human-escalation patterns in 7 of the 50 analysed dialogues. Booking-intent conversion is consistent across the two principal clinics (38.7% and 39.4%), although their overall Success rates differ in the unadjusted comparison, a gap

that is mostly a calendar-time effect and narrows from 29.9 to 11.8 percentage points once both clinics are restricted to the months they share (Sections 4.6 and 4.8), and the findings align with conceptual expectations drawn from the published literature. These aggregates span a period of active development: in the final three months of the window the system reached 86.2% Success and 32.8% booking completion, against 26.2% and 3.3% in the first three (Section 4.8).

For dental practices and technology providers, the framework offers a set of design and governance principles anchored to a real-world operational profile. It supports incremental adoption, starting with self-service booking and FAQs and progressing toward integrated triage of routine requests and data-driven scheduling proposals. The empirical observation that booking-intent conversion converges to ≈ 39% across deployments while overall conversion varies widely shows that the principal lever for performance gains lies upstream, in intent capture, opening-turn design and reduction of short abandoned interactions, and not in the scheduling logic itself.

Future empirical work can build on this conceptual and empirical foundation through controlled before-and-after comparisons in clinics with detailed historical call logs, evaluation of multilingual and multichannel performance in linguistically diverse populations, and quantification of staff-time displacement and patient-experience effects through primary survey instruments. Targeted improvements to operational telemetry, including populating the escalation, patient-status, and language fields, would enable deeper analyses than the current data layer permits.

Viewed through an operations management lens, the virtual receptionist serves as an optimisation-driven control layer atop existing practice management infrastructure. Data from intake, triage and historical attendance feed into appointment-allocation models that smooth demand, reduce idle time and compress avoidable waiting. Progressive deployment of such systems offers a path toward measurable gains in workflow efficiency and toward data-informed resource allocation in dental practice operations.

CONFLICT OF INTEREST

The author is the CEO and Co-Founder of DentalAssist.ai, the commercial system evaluated in this study. The analysis was conducted on anonymised operational data and all findings are reported in aggregate form. This affiliation is disclosed in the interest of transparency.

FUNDING

This study received no external funding.

REFERENCES

[1] Barreda, M., Cantarero-Prieto, D., Coca, D., Delgado, A., Lanza-León, P., Lera, J., Montalbán, R., & Pérez, F. (2025). Transforming healthcare with chatbots: Uses and applications - A scoping review. *Digital Health, 11*, 20552076251319174. https://doi.org/10.1177/20552076251319174

[2] Laymouna, M., Ma, Y., Lessard, D., Schuster, T., Engler, K., & Lebouché, B. (2024). Roles, users, benefits, and limitations of chatbots in health care: Rapid review. *Journal of Medical Internet Research, 26*, e56930. https://doi.org/10.2196/56930

[3] Grassini, E., Buzzi, M., Leporini, B., & Vozna, A. (2025). A systematic review of chatbots in inclusive healthcare: Insights from the last 5 years. *Universal Access in the Information Society, 24*, 195–203. https://doi.org/10.1007/s10209-024-01118-x

[4] Knight, R. T. D., Aakre, C. A., Anstine, C. V., Munipalli, B., Biazar, P., Mitri, G., Valery, J. R., Brigham, T., Niazi, S. K., Perlman, A. I., Halamka, J. D., & Abu Dabrh, A. M. (2023). Artificial intelligence for patient scheduling in the real-world health care setting: A metanarrative review. *Health Policy and Technology, 12*(4), 100824. https://doi.org/10.1016/j.hlpt.2023.100824

[5] Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: A review of literature. *Production and Operations Management, 12*(4), 519–549. https://doi.org/10.1111/j.1937-5956.2003.tb00218.x

[6] Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions, 40*(9), 800–819. https://doi.org/10.1080/07408170802165880

[7] Green, L. (2006). Queueing Analysis in Healthcare. In: Hall, R.W. (eds) *Patient Flow: Reducing Delay in Healthcare Delivery*. International Series in Operations Research & Management Science, vol 91. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-33636-7_10

[8] Hulshof, P. J. H., Kortbeek, N., Boucherie, R. J., Hans, E. W., & Bakker, P. J. M. (2012). Taxonomic classification of planning decisions in health care: A structured review of the operations management literature. *Health Systems, 1*(2), 129–175. https://doi.org/10.1057/hs.2012.18

[9] Alsalamah, S. A., AlSalamah, S., Alsalamah, H. A., Sheerah, H. A., Luther, K., & Lu, C.-T. (2025). Virtual healthcare bot (VHC-Bot): A person-centred AI chatbot for transforming patient care and healthcare workforce dynamics. *Netw Model Anal Health Inform Bioinforma, 14*, 48. https://doi.org/10.1007/s13721-025-00537-x

[10] Esmailpour, H., Rasaie, V., Babaee Hemmati, Y., & Falahchai, M. (2025). Performance of artificial intelligence chatbots in responding to the frequently asked questions of patients regarding dental prostheses. *BMC Oral Health, 25*, 574. https://doi.org/10.1186/s12903-025-05965-9

[11] Maruska, E., Elgreatly, A., Madaio, W., Razoky, K., Bay, C., & Mahrous, A. (2025). Comparing dentist and chatbot answers to dental questions for quality and empathy. *JADA Foundational Science, 4*, 100044. https://doi.org/10.1016/j.jfscie.2025.100044

[12] Suárez, A., Adanero, A., Díaz-Flores García, V., Freire, Y., & Algar, J. (2022). Using a virtual patient via an artificial intelligence chatbot to develop dental students' diagnostic skills. *International Journal of Environmental Research and Public Health, 19*(14), 8735. https://doi.org/10.3390/ijerph19148735

[13] Shaikh, U., Mustafa, S., Mujawar, S., Shaikh, H., Pathan, Z., & Shaikh, J. (2025). AI healthcare chatbot: Enhancing healthcare delivery for patients. *International Journal of Science and Management Studies, 8*(3), 243–253. https://doi.org/10.51386/25815946/ijsms-v8i3p120

[14] Geoghegan, L., Scarborough, A., Wormald, J. C. R., Harrison, C. J., Collins, D., Gardiner, M., Bruce, J., & Rodrigues, J. N. (2021). Automated conversational agents for post-intervention follow-up: A systematic review. *BJS Open, 5*(4), zrab070. https://doi.org/10.1093/bjsopen/zrab070