Proposal

Title: Evaluation of a Chatbot Using Large Language Models for Patient Care after Oral Surgery

หัวข้อเรื่อง: การประเมินประสิทธิภาพของแชทบอทที่ใช้แบบจำลองภาษาขนาดใหญ่ในการดูแลผู้ป่วยหลังการผ่าตัดในช่องปาก

Investigator: Miss Nopasorn Wilairatanaporn 6430038332

Advisors: Assist. Prof. Vorapat Trachoo, Department of Oral and Maxillofacial Surgery

Type of research: Applied research

Field of research: Miscellaneous (Artificial Intelligence)

Keywords: Large language models, Artificial intelligence, Oral surgery, Postoperative care, Patient care, Chatbot

Chapter1

Background and Rationale

Artificial intelligence (AI) represents one of the most transformative technological advancements of the modern era. It encompasses various technologies that enable machines to simulate human intelligence, performing tasks traditionally requiring human cognition, such as learning, problem-solving, and decision-making. In recent years, AI has been applied in a wide range of specialties and opened a wealth of exciting opportunities. A major development has been made in dentistry; AI has emerged as a transformative tool with diverse applications across different domains, including diagnosis, treatment planning, and patient management (1, 2). Particularly in oral surgery, the postoperative period is critical for initiating healing and preventing surgery-related complications (3). Traditional postoperative follow-up via telephone has been proven effective, offering benefits such as reduced consultation costs compared to conventional on-site follow-ups. However, in clinical practice, telephone follow-ups also present significant challenges. Physicians and nurses often spend considerable time repeating the same instructions or guidelines, and patients often reject the calls, as they think that these unknown incoming calls are harassing calls (4). This issue is further exacerbated by the increasing prevalence of scam calls. In 2021 alone, Thailand recorded over 6.4 million scam calls from call center gangs, marking a 270% increase from the previous year (5). These challenges highlight the need for innovative solutions to streamline post-operative care while maintaining effective patient-provider communication.

One promising solution lies in the integration of AI-driven chatbots into dental care. Chatbots are sophisticated conversational programs designed to simulate human interaction (6). The first well-established chatbot of this kind, ELIZA, was programmed in 1966 to simulate a text-based conversation with a psychotherapist and has had a remarkable impact on how humans interact with machines and seek answers (7). Significant advances in AI algorithms have led to the development of large language model (LLM)-based chatbots, AI chatbots that leverage large language models (LLMs) to generate coherent and contextually relevant text based on prompts. For example, ChatGPT has experienced rapid growth and has significantly impacted communication and information access (8). As healthcare gravitates towards digitalization and AI-driven processes (6), especially in light of the COVID-19 pandemic and rising concerns over environmental factors like PM2.5 air pollution, the adoption of chatbots in dentistry has gained momentum (9), particularly in patient care after oral surgery, driven by their ability to understand and interpret natural language input from patients, analyze queries, and retrieve relevant information from their knowledge bases. They can provide accurate, timely, and personalized responses to postoperative patient's queries (10). They can also monitor patients remotely following oral surgery by checking in with patients to assess their recovery progress, asking about pain levels, medication adherence, and any unexpected symptoms to classify postoperative complication risk levels (Low/Moderate/High), and offer tailored recommendations based on their condition (1). This proactive approach enables early identification of complications or infections, potentially leading to faster and more effective interventions (11, 12). Moreover, AI chatbots can provide patients with round-the-clock access to post-operative consultations, reducing unnecessary and unplanned hospital visits. Many common and expected postoperative concerns can be resolved through explanations and reassurance, eliminating the need for in-person consultations (4). This not only alleviates the workload of healthcare providers but also allows them to focus on patients who require more urgent attention. (13) A non-randomized comparative study in orthopedic surgery demonstrated that an AI-assisted follow-up conversational agent achieved comparable effectiveness with manual telephone-based follow-up while saving an estimated 9.3 hours per 100 participants (14), underscoring the potential efficiency gains of such systems.

This study aims to comprehensively evaluate a newly developed chatbot using LLMs for patient care after oral surgery. The evaluation will focus on assessing the chatbot's performance and its impact on healthcare providers' workflows, including its ability to accurately and effectively evaluate postoperative oral surgery complication risk and response to patients's inquiries. The chatbot will be developed to serve as a tool to monitor patients' symptoms, classifypostoperative complication risk levels (low/moderate/high), provide personalized recommendations tailored to their postoperative conditions and concerns, and alert healthcare providers if there are any concerning symptoms. This evaluation will examine how effectively the chatbot performs these functions compared to traditional telephone-based follow-up methods, with particular attention to healthcare provider time savings, system acceptability, and ease of use.

To ensure balanced integration of technology with traditional healthcare practices, dental professionals will oversee the chatbot's responses, manage complex cases that require direct intervention, and validate the information provided, ensuring that surgeons' expertise and clinical judgment remain irreplaceable (14).

The anticipated outcomes of this evaluation study include measurable metrics on the chatbot's accuracy in evaluating postoperative oral surgery complication risk, patient interactions and management, quantifiable improvements in healthcare provider workflow efficiency, and insights into user satisfaction. These findings could demonstrate enhanced quality of care, reduced workload for healthcare providers, and ultimately lead to better patient outcomes. Furthermore, this research seeks to contribute valuable insights into how technology can complement conventional healthcare practices. By rigorously evaluating an AI-driven solution in oral surgery, this study aims to provide evidence-based recommendations for future innovations in digital dentistry, particularly within Thailand's rapidly evolving healthcare landscape.

Chapter2

Literature review

Artificial intelligence (AI) can be defined as a sequence of operations designed to perform specific tasks. It is a branch of computer science focused on understanding and creating intelligent entities capable of simulating human intelligence (15). These systems are designed to perform tasks that typically require human cognition, such as learning, reasoning, problem-solving, perception, and decision-making (1). For decades, scientists have worked tirelessly to advance this field, which was formally introduced by John McCarthy in 1956. It is, at times, called machine intelligence. AI is also considered a cornerstone of the "fourth industrial revolution," employing computer technology to mimic critical thinking, decision-making, and intelligent behavior similar to that of humans (16). The core component of artificial intelligence technology is a neural network, which is designed to mimic the structure of the human brain and simulate human thought. These networks consist of strongly interconnected neurons that function as data processing systems to address specific problems. As a rapidly evolving technology, AI enables machines to perform tasks that are formerly human-only jobs (16).

Advances in AI have led to the development of conversational agents (chatbots), which are computer programs designed to simulate human interactions. These chatbots employ intricate algorithms, machine learning methodologies, and natural language processing (NLP), to interpret and respond to human queries. NLP allows computers to understand and interpret human language, human speech subtleties, and generate human-like responses (6). The first known agent capable of conversation between human and machine, Eliza, was developed in 1966 using early NLP techniques to return open-ended questions to users, simulating person-centered psychotherapy (17). Recent advancements in generative AI and large language models (LLMs) have significantly propelled the evolution of intelligent chatbots . By ingesting and processing vast amounts of data, LLMs have developed the ability to understand human language and generate responses that closely mimic human interactions (6). These models can create original content based on their training data, determine the most probable response to a given input, and provide real-time answers across a wide range of applications from creative tasks to solving complex problems (18).

The introduction of LLM-based chatbots such as ChatGPT (Chat Generative Pre-trained Transformer; OpenAI, San Francisco, United States) and Bard (Language Model for Dialogue Applications; Google, Mountain View, United States) has marked a significant milestone in the application of AI in medicine. These chatbots have demonstrated the ability to provide medical knowledge comparable to that of a third-year medical student or first-year medical resident (8). For instance, ChatGPT’s success in passing the United States Medical Licensing Exam (USMLE) without specialized training garnered global attention in January 2023 (19). While the use of chatbots in healthcare is still relatively new, acceptability and patient experience have so far been largely positive (20). However, the adoption of AI-driven chatbots in dentistry has lagged significantly behind their counterparts in medicine. It was only towards the end of 2023 that reports on the potential of chatbots in dentistry began to emerge (19).

In dentistry, chatbots serve as valuable tools for dental self-diagnosis, information dissemination, and remote consultations, reducing the need for in-person visits to dental clinics (9). This is particularly beneficial in patient care after oral surgery, where the postoperative period is critical for initiating healing and preventing complications that could cause discomfort or negatively impact a patient’s quality of life, such as nerve damage, infection, hemorrhage, exacerbation of pain, and surgical relapse (3, 4, 21). In addition, patients often struggle with compliance, comprehension, and retention of postoperative care instructions, leading to suboptimal recovery outcomes (3).

The integration of LLM-based chatbots into patient care after oral surgery presents a promising opportunity to enhance patient experience and outcomes (17). Studies have already demonstrated that chatbots can significantly improve postoperative experiences, such as in hip arthroscopy patients, where high levels of satisfaction were reported (20). Additionally, mobile applications delivering postoperative care have shown benefits, including reduced in-person follow-ups (20), decreased opioid prescribing (14), and significant time and cost savings. These findings provide compelling evidence supporting the investment in automated follow-up systems (17).

Despite these advancements, the field remains in its early stages. By 2024, reputable journals were primarily publishing pilot studies exploring the application of general LLM-based chatbots, such as Google Bard and ChatGPT, in dentistry. These studies highlighted significant limitations, including the chatbots' poor performance in effective communication compared to human dental specialists (22). Furthermore, LLMs like ChatGPT are trained with fixed information cutoffs (e.g., September 2021 for GPT-4), which limits their ability to provide up-to-date scientific information (23).

A potential solution to this limitation is the use of Retrieval Augmented Generation (RAG), a technique that enhances LLMs' performance by dynamically retrieving relevant information from external knowledge bases. The RAG approach allows the system to pull relevant documents from its knowledge base and provide precise answers to complex dental queries (22). It also enables simple modification to incorporate evolving scientific knowledge, as its architecture enables dynamic updating of the external knowledge source. By regularly updating the knowledge retrieval component with the latest scientific literature, the model can easily be adapted to new findings and research. This flexibility makes RAG an attractive option for tasks requiring up-to-date information. Additionally, RAG allows users to explore topics in greater depth by accessing referenced sources via hyperlinks, a feature particularly valuable when dealing with conflicting information or complex answers (23).

To further improve the system's efficiency, an embedding model can be used to transform the knowledge database into a dense vector representation stored in a vector database. This approach organizes information in a structured manner, positioning similar texts closer together in the vector space (22, 23). As a result, the LLM can search and produce relevant outputs more effectively in response to user queries (1, 24).

While the utilization of LLM-based chatbots in clinical settings holds tremendous potential for enhancing patient care after oral surgery, it also raises several ethical considerations. First, issues related to patient autonomy and informed consent must be addressed. Patients should be adequately informed about the role of AI in their care and have the right to understand and consent to AI-driven interventions, with the option to opt-out if desired (1, 16). Second, the transparency and interpretability of AI algorithms need to be considered. Many AI models are often perceived as "black boxes" due to their complex decision-making processes, posing challenges for accountability (25). Developing transparent AI systems, establishing clear guidelines, and the involvement of competent clinicians for accountability are essential steps to address this issue (1, 26, 27). Third, bias in AI algorithms is another critical ethical concern. AI models trained on biased or incomplete datasets may inadvertently reinforce existing healthcare disparities, leading to partial or irrelevant responses. To mitigate bias and prevent discrimination in patient care, it is necessary to ensure diverse and representative training data, rigorous algorithm testing, and ongoing monitoring (1, 6, 27). Finally, privacy and data security are paramount. AI systems rely heavily on sensitive patient data for training algorithms and making accurate predictions, making it imperative to implement robust data governance frameworks, stringent access controls, and anonymization techniques to safeguard the privacy and security of these data and prevent potential breaches, thereby upholding patient trust (1, 8, 27).

In conclusion, the increasing adoption of AI in dentistry is accompanied by ethical concerns that must be addressed to ensure responsible integration. Before incorporating AI models into routine clinical operations, it is essential to certify their accuracy, cost-effectiveness, dependability, and applicability (1, 2, 28). There is also a risk of overreliance on technology, which could diminish the human element in dental care and impact patient-provider relationships (1). Therefore, maintaining a balanced approach is crucial. AI systems, while powerful, still require oversight from surgeons and healthcare teams (1, 24, 27). The complexity of these challenges underscores the need for comprehensive strategies that address technological, ethical, and regulatory dimensions. By doing so, we can maximize the benefits of AI while ensuring its responsible and secure integration into healthcare environments (29).

Research questions

1. Can an LLM-based chatbot accurately assess postoperative symptoms and classify complication risk levels (low/moderate/high) from nurse-documented telephone records?
2. How clinically appropriate are the chatbot's generated recommendations (for low/moderate-risk cases) and alerts (for high-risk cases) for postoperative oral surgery care, as judged by healthcare providers?
3. Does the chatbot significantly reduce healthcare providers' time spent on postoperative follow-ups and/or improve their satisfaction compared to telephone-based methods?

Research objectives

This study aims to evaluate the performance and impact of an LLM-based chatbot for postoperative oral surgery care by

1. Assessing accuracy in classifying patient-reported symptoms into low-, moderate-, and high-risk categories based on past medical records.

2. Evaluating the chatbot’s ability to generate tailored postoperative care recommendations for low- and moderate-risk cases or provide timely alerts to healthcare providers for high-risk cases. (Metrics: response accuracy, relevance, completeness, and conciseness.)
3. Measuring healthcare provider satisfaction with the chatbot’s summarized reports and its impact on clinical workflow efficiency.

Research hypothesis

H_{01}: The LLM-based chatbot's accuracy in assessing postoperative symptoms and classifying postoperative complication risk levels (low/moderate/high) based on nurse-documented telephone follow-up records shows less than 80% agreement with the actual clinical outcomes recorded in patient files and the management decisions documented.

H_{11}: The LLM-based chatbot's accuracy in assessing postoperative symptoms and classifying postoperative complication risk levels (low/moderate/high) based on nurse-documented telephone follow-up records shows more than 80% agreement with the actual clinical outcomes recorded in patient files and the management decisions documented.

H_{02}: The chatbot's responses (recommendations/alerts) do not meet minimum clinical standards for accuracy, relevance, completeness, and conciseness.

H_{12}: The chatbot's responses meet or exceed minimum clinical standards for accuracy, relevance, completeness, and conciseness.

H_{03}: The chatbot does not significantly reduce healthcare providers’ time spent on follow-ups or improve their satisfaction compared to telephone-based methods.

H_{13}: The chatbot significantly reduces time spent on follow-ups and/or improves provider satisfaction compared to telephone-based methods.

Benefits of the study

This study will contribute to the growing field of AI-driven healthcare solutions by providing an evidence-based evaluation of a chatbot using Large Language Models (LLMs) for postoperative oral surgery care. By rigorously assessing the chatbot’s performance, this research will determine its effectiveness in assessing postoperative symptoms, classifying postoperative complication risk levels (low/moderate/high), delivering accurate, appropriate, and tailored recommendations to patients (for low/moderate-risk cases) or alerting healthcare providers (for high-risk cases), and its impact on healthcare providers’ workflows, particularly in terms of time savings and satisfaction compared to traditional telephone-based follow-up methods.

The findings from this study could lead to the development and enhancement of the chatbot's capabilities, expected to provide timely responses, even outside clinic hours, ensuring that patient concerns are addressed promptly and providing guidance on whether symptoms require urgent medical attention. Minimizing unnecessary hospital visits and lowering costs for both patients and the healthcare system. Additionally, the chatbot may help identify early signs of complications, reducing the risk of adverse effects and improving patient outcomes. By automating routine follow-ups and postoperative symptom management recommendations, the chatbot will reduce the workload of healthcare providers, allowing them to focus on more critical tasks and enabling hospitals to allocate staff and equipment more effectively. This, in turn, will enhance overall service quality. Furthermore, the chatbot's ability to automate and standardize documentation will improve the accuracy and completeness of patient records, supporting more informed clinical decision-making and enhancing the quality of postoperative monitoring. By enabling patients to revisit chatbot-provided guidance at any time, the system may also improve patient compliance with postoperative care recommendations, ultimately leading to better recovery outcomes and paving the way for future innovations in digital dentistry.

Looking toward future applications, this study will identify pathways for expanding the chatbot to other surgical specialties and establish a framework for integration with electronic health records. The research will develop standards for secure, HIPAA-compliant AI implementation in healthcare and contribute to evidence supporting regulatory approvals for AI-based medical support tools, ultimately paving the way for future innovations in digital dentistry.

Chapter3

Research Methodology

The research methodology for this study focuses on the evaluation of a newly developed Chatbot Using Large Language Models for Patient care after oral surgery. The methodology is divided into three phases: model development, system integration, chatbot evaluation.

Model development

To develop a new chatbot using LLMs tailored for patient care after oral surgery, the project aims to effectively monitor patients' symptoms and predict complication risks, offering personalized recommendations to patients (for low/moderate-risk cases) or alerting healthcare providers (for high-risk cases), thereby alleviating the workload of healthcare providers. The development process is organized into several key phases: data collection, data processing and knowledge base construction, and model training.

1. Data Collection

To ensure the chatbot is effective and relevant, data collection is a critical step. The following methods will be used to gather the necessary information:

1.1 Interviews with Healthcare Providers

To gain invaluable insights from dentists, nurses, and call-center representatives who handle postoperative follow-ups by conducting an on-site observation and in-depth interviews at the Oral surgery department in the Faculty of Dentistry, Chulalongkorn University. The participants include dentists who Responsible for routine on-site follow-ups, Nurses who responsible for inpatient department (IPD) follow-ups three days after a patient's discharge, and call-center representatives who employed at the outpatient department (OPD) Instant Clinic's call center, responsible for OPD patient follow-ups one day after their surgery.

The interview questions will cover patient symptom monitoring, typical responses and recommendations provided by healthcare providers, differences in symptom management among various procedures, common challenges faced during postoperative follow-ups, Feedback on the current telephone-based follow-up system, including its limitations and areas for improvement.

1.2 Patient Data Collection

Patient data will be collected from past patient records and discharge notes that were documented by nurses during telephone-based follow-up sessions three days after discharge. These records include essential details such as the operation performed, discharge date, follow-up date, symptoms observed, additional notes, and management given. This rich dataset of patient interactions, symptoms, and management will be anonymized and prepared for training and evaluating the chatbot.

1.3 Knowledge Base Construction

To incorporate evidence-based guidelines into the chatbot's knowledge base (vector DB), PDF files containing clinical guidelines and examples of postoperative complication cases with appropriate management for patient care after oral surgery will be collected from dental associations, professional organizations, and research articles. These documents will be classified into three reliability levels: low, medium, and high, and only medium and high reliability will be used. This structured knowledge base will enable the chatbot to provide accurate and standardized responses.

2. Data Processing to Construct a Knowledge Base

Once the data is collected, it will be processed to ensure it is suitable for training and integration into the chatbot system. This study employs a semi-structured document chunking approach, which includes document preprocessing, recursive character splitting, semantic splitting, and metadata enrichment to enhance efficiency while preserving the contextual meaning of the content.

2.1 Document Preprocessing

Data will be imported from PDF files and cleaned using Unicode Normalization to convert special characters and Thai text for correct display, White Space Normalization to standardize spacing and line breaks, Metadata Management to track document sources and text locations for easy reference.

2.2 Recursive Character Splitting

Documents will be split into smaller chunks (paragraphs, sentences, words) using a hierarchical text-splitting approach. The chunk size and overlap will be defined to maintain contextual integrity, and the structural accuracy of the segmented text will be validated.

2.3 Semantic Splitting

Embedding models will be used to analyze semantic relationships within the content. These models convert text into vector representations, allowing for the computation of similarity scores using cosine similarity. Chunks will be merged or restructured based on similarity scores to ensure semantic completeness. Four efficient Thai embedding models will be compared for semantic search:

- Cohere-embed-multilingual-v2.0
- SCT-Distil-model-phayathaibert-bge-m3
- ConGen-BGE_M3-model-phayathaibert
- 3b-bge-m3

The performance of these models will be evaluated using Semantic Similarity Evaluation, which measures the semantic similarity between queries and retrieved documents. A higher similarity score indicates a better ability to match semantically related content.

2.4 Metadata Enrichment

Metadata, including document ID, section titles, and positional indices, will be assigned to each chunk to enhance retrieval efficiency in search and indexing processes.

3. Model Training

Three Thai-optimized LLMs (Llama-3-8B, Typhoon2-Qwen2.5-7B, and Qwen2.5-7B) will be fine-tuned using historical nurse documentation of telephone follow-ups. The training data pairs symptom descriptions with gold-standard risk classifications (Low/Moderate/High) established by surgeon review. Parameter-efficient fine-tuning (LoRA) preserves base model capabilities while adapting to clinical terminology.

The chatbot system will utilize LangChain, which enables seamless integration between a vector database and an LLM. The RAG approach will be implemented to enhance the chatbot's ability to generate accurate and context-aware responses (recommendations/alerts) by incorporating external knowledge sources.

The system follows a structured process:

1. <u>User Query Processing:</u> The user's query is converted into an embedding representation using an embedding model.
2. <u>Context Retrieval:</u> The system searches the vector database for semantically relevant documents and retrieves the top three most relevant ones.
3. <u>Prompt Augmentation:</u> The retrieved documents are combined with the original query to form a context-enhanced prompt.
4. <u>Response Generation:</u> The augmented prompt is processed by the LLM to generate a response that is more accurate and contextually relevant.

4. System Integration

The chatbot will be deployed through a Line Official Account, providing an accessible interface for both patients and healthcare providers. The system will be developed using Streamlit for backend logic and integrated with the Line Messaging API for real-time interaction. The chatbot interface will be developed using Streamlit, a Python-based framework for building Line official chatbots. The interface will include two main features:

1. <u>Patient-Facing Chat Interface</u>

   For patients, the chatbot conducts structured symptom assessments via a dynamic decision tree, classifying risk levels (low/moderate/high) and delivering tailored guidance—from self-care advice to urgent visit recommendations.

   Sample questions:

1. General symptoms
- How is the patient doing?
(อาการทั่วไป - ผู้ป่วยเป็นอย่างไรบ้าง?)
2. Bleeding
- Is there still oozing blood?
(ยังมีเลือดซึมอยู่หรือไม่?)
3. Oral hygiene
- Can the patient brush their teeth and rinse their mouth? Does the patient have mouthwash or saline solution for rinsing?
(สามารถแปรงฟัน บ้วนปากได้หรือไม่ ผู้ป่วยมีการใบ้น้ำยาบ้วนปากหรือน้ำเกลือบ้วนปาก?)
4. Food intake
- For patients who did Intermaxillary Fixation (IMF) with wires: Are they still using a syringe to assist with eating, consuming soft liquid foods, avoiding spicy/strong-flavored foods?
(กรณีมีการมัดฟัน: ผู้ป่วยยังใช้ syringe ช่วยทานอาหาร, ทานอาหารเหลวอ่อน, งดอาหารรสจัดอยู่หรือไม่?)

- For patients with nasogastric tube: Are the tube and nasal tape still remains in position?
(กรณีผู้ป่วยมีการให้อาหารผ่านสายยาง: สายยางและเทปยึดจมูกยังอยู่ในตำแหน่งหรือไม่?)
5. Signs of infection
- Does the patient have high fever, pain, swelling, redness, heat, or foul-smelling discharge from the surgical wound?
(ผู้ป่วยมีอาการไข้สูง ปวด บวม แดง ร้อน มีสารคัดหลั่งกลิ่นเหม็นไหลซึมจากแผลผ่าตัดหรือไม่?)
6. Phlebitis (vein inflammation)
- At the Intravenous needle removal site (back of hand or wrist), is there pain, swelling, redness, or red streaks along the vein?
(บริเวณที่เอาเข็มน้ำเกลือออก หลังมือหรือข้อมือมีอาการปวดบวมแดง หรือพบรอยแดงเข็มตามเส้นเลือดหรือไม่?)
7. Cold/warm compress
- If there is oozing blood or within 1-3 days after surgery, is the patient applying a cold compress?
(หากมีเลือดซึมหรือ 1-3 วันหลังผ่าตัด ผู้ป่วยมีการประคบเย็นอยู่หรือไม่?)
-If there is facial swelling or 4-6 days after surgery, is the patient applying a warm compress?
(หากมีหน้าบวมหรือ 4-6 วันหลังผ่าตัด ผู้ป่วยมีการประคบอุ่นอยู่หรือไม่?)
8. Medication
- Has the patient completed the full course of antibiotics as prescribed?

(ผู้ป่วยมีการรับประทานยาฆ่าเชื้อครบโดสตามที่ให้ไปหรือไม่?)

- If experiencing pain, did the patient receive ibuprofen or paracetamol?

(หากมีอาการปวดผู้ป่วยได้รับยา ibuprofen หรือ paracetamol หรือไม่?)

9. For patients discharged with IMF

- For those who did IMF with wires: Are the wires loose? Can the patient open their mouth?

(ในรายที่มัดลวด ลวดมัดฟันหลวมหรือไม่ สามารถอ้าปากได้หรือไม่?)

- For those who did IMF with elastic bands: Can the patient remove the elastics bands?

(ในรายที่มัดยางสามารถถอดยางมัดฟันได้หรือไม่?)

2. Healthcare Provider Dashboard

The system generates summarized reports of patient interactions, highlighting risk levels and chatbot recommendations. High-risk cases trigger real-time alerts, enabling prompt clinical intervention.

This integration optimizes clinical efficiency while maintaining patient accessibility, bridging AI-driven triage with real-world care workflows.

Evaluation

To rigorously test the study hypotheses while incorporating both clinical and technical perspectives, we implement a hybrid evaluation framework combining expert assessment with automated metrics:

1. **Chatbot's performance evaluation**

1.1 Clinical Accuracy Validation

The primary evaluation will assess the chatbot's classification accuracy against clinical outcomes. A retrospective dataset of 300 nurse-documented telephone follow-up records will be processed to establish gold-standard labels based on actual clinical outcomes and appropriate management decisions. The chatbot's risk assessments will be compared against these standards. The 80% accuracy threshold will be statistically tested using binomial analysis with α=0.05.

1.2 Recommendation Quality Assessment

We combine expert human judgment with automated natural language processing metrics. A panel of five clinicians will assess 30 responses across four dimensions: accuracy, relevance, completeness, and clarity, using validated 5-point Likert scales. This qualitative assessment is complemented by quantitative text analysis, including BERTScore for semantic similarity, BLEU/METEOR/ROUGE-1 for generation quality, and Flesch readability metrics to ensure health literacy appropriateness (targeting a Flesch Reading Ease (FRE) score ≥80 and a Flesch–Kincaid Grade Level (FKGL) ≤6.0).

The FRE and FKGL scores will be particularly important to ensure responses meet healthcare readability standards. The FRE score ranges from 1 to 100, with higher scores indicating easier readability. The FKGL, on the other hand, estimates the required reading grade level for comprehension. Patient-facing materials should not exceed the 6th- to 8th-grade reading levels, according to AMA and NIH guidelines (13).

2. Comparative Assessment with Traditional Methods

To assess the chatbot's impact compared to traditional telephone-based follow-ups, an in-depth survey and interviews will be conducted with healthcare providers. Dentists, nurses, and OPD call-center representatives will complete a structured questionnaire evaluating both chatbot-based and telephone-based follow-ups across key dimensions:

1. Time Efficiency: Time spent on follow-ups with each method (measured in minutes)
   Sample questions:
   1. On average, how much time do you spend per patient follow-up using [chatbot/telephone]?
   -Less than 2 minutes/ 2–5 minutes/ 5–10 minutes/ More than 10 minutes
   2. Can you describe how the [chatbot/telephone] follow-up process affects your daily workflow? Does it save time or create additional burdens?
2. Satisfaction: Provider satisfaction with each method (Likert scale: 1 = strongly disagree, 5 = strongly agree)
   Sample questions:
   1. How satisfied are you with using [chatbot/telephone] for post-operative follow-ups?
   2. What do you see as the biggest advantages and drawbacks of using [chatbot/telephone] for post-operative follow-ups?
   3. Have you encountered situations where the chatbot missed critical patient concerns that a phone call might have caught? Can you share an example?
3. Ease of Use: Perceived usability of each method (Likert scale: 1 = very difficult, 5 = very easy)
   Sample questions:
   1. How easy is it to use the [chatbot/telephone] system for patient follow-ups?
   2. What technical or logistical challenges have you faced while using the [chatbot/telephone] system?
   3. How intuitive is the chatbot interface compared to making phone calls? What improvements would you suggest?

4. Clinical Efficacy: Perceived effectiveness in identifying patient issues
   Sample questions:
   1. How effective is the [chatbot/telephone] in identifying patients who require urgent clinical attention?
   2. In your experience, how reliable is the chatbot in escalating high-risk patients? Are there specific cases where it failed or excelled?
   3. Do you trust the chatbot's ability to triage patients as effectively as a human phone conversation? Why or why not?
5. Documentation Quality: Completeness and accuracy of patient interaction records
   Sample questions:
   1. How would you rate the completeness and accuracy of patient records generated by [chatbot/telephone] follow-ups?
   2. How useful are the chatbot-generated records for clinical decision-making? Do they provide sufficient detail compared to phone notes?
   3. Have you noticed any gaps or errors in the chatbot's documentation of patient interactions?"

This comparative assessment will provide quantitative and qualitative insights into the relative advantages and limitations of each approach.

3. **Patient Usability Evaluation**

   This study assesses patient experiences with the chatbot through structured testing with 30 postoperative patients. The evaluation uses a mixed-methods approach:
   1. Quantitative Assessment: Patients complete a 5-point Likert scale survey evaluating ease of use (navigation, interface clarity), response quality (accuracy, helpfulness), and overall satisfaction.
   2. Qualitative Feedback using brief interviews explores preferred features, improvement suggestions, and comparisons to phone follow-ups.

Sample Size Calculation

This study will evaluate the chatbot system across four key domains: classification accuracy, recommendation quality, workflow efficiency, and patient usability evaluation.

To determine the required sample size for the primary accuracy validation, power analysis will be conducted based on a one-sample binomial test. With an α of 0.05 and 80% power, to calculate that 200 cases would reliably test whether the chatbot meets the 80% accuracy threshold, detecting a minimum 5% improvement (85% vs 80%) with 80% power.

For recommendation quality evaluation, applied inter-rater reliability analysis (target ICC>0.7) to determine that 30 responses (20 low/moderate-risk, 10 high-risk) would provide sufficient precision when evaluated by five clinicians. This allows detection of 1-point Likert scale differences (SD=1.5) with 90% power using paired t-tests.

For workflow analysis among healthcare providers (dentists, nurses, and OPD call-center representatives) in patient care after oral surgery, evaluating chatbot-based follow-ups versus telephone-based follow-ups using Linkert scale data. To determine the required sample size power analysis will be conducted. Given that the same participants will rate both methods, a paired sample t-test (or Wilcoxon Signed-Rank test if the data is non-normal) will be used for hypothesis testing. Assuming a moderate effect size (0.5), a significance level of 0.05, and a power of 80%, the estimated minimum sample size is 34 participants. To account for potential dropouts or incomplete responses, this will be increased by 10–20%, resulting in a target sample size of 40 participants.

Finally, a patient usability evaluation that evaluates the chatbot through a mixed-methods approach with 30 postoperative patients, combining quantitative surveys and qualitative interviews. The sample size provides 80% power (α=0.05) to detect ≥1-point satisfaction differences (effect size=0.7). Following FDA guidelines, the evaluation occurs in two phases: formative testing (n=15) to refine the interface and validation (n=15) assessing usability via 5-point Likert scales and thematic interviews. Quantitative analysis uses Wilcoxon tests for Likert data; qualitative analysis continues until saturation. The design includes FDA-required safeguards: clinician oversight for high-risk cases, documented failure modes, TLS-encrypted Line API communications, and PDPA-compliant anonymization. This approach balances statistical rigor with regulatory compliance for clinical implementation.

Ethical Considerations

The study will address several ethical considerations, including:

- Patient Autonomy and Informed Consent: Patients will be informed about the chatbot's role in their care and will have the option to opt-out.
- Transparency and Interpretability: Transparent AI systems will be developed, and clear guidelines for accountability will be established.
- Bias Mitigation: Diverse and representative training data will be used to prevent bias in the chatbot's responses.
- Privacy and Data Security: Robust data governance frameworks will be implemented to safeguard patient data.

Knowledge Gap and Limitations

While LLMs have shown promise in various domains, their application in patient care after oral surgery remains underexplored. Existing chatbots in healthcare often lack the ability to provide context-aware, personalized responses, especially in languages other than English. This study aims to bridge this gap by developing a Thai-language chatbot that leverages LLMs and RAG to provide accurate, context-aware responses tailored to patient care after oral surgery.

However, the study has several limitations. The chatbot is primarily designed for Thai-speaking patients, which may limit its applicability in multilingual settings. The quality of the chatbot's responses depends on the availability and quality of the training data, which may be limited in certain areas of postoperative care. While LLMs are powerful, they may still generate inaccurate or irrelevant responses, especially when dealing with rare or complex cases, necessitating continuous evaluation and fine-tuning. Ensuring patient privacy and data security while using AI in healthcare remains a significant challenge, requiring ongoing attention and robust safeguards.

By addressing these gaps and limitations, this study aims to contribute to the growing field of AI-driven healthcare solutions, particularly in patient care after oral surgery.

Statistical analysis

This study employs a comprehensive statistical approach to evaluate the chatbot's performance across three key domains: clinical accuracy, recommendation quality, and workflow efficiency. The analysis incorporates both quantitative and qualitative methods to provide robust evidence for the chatbot's effectiveness.

For clinical accuracy validation, we will use a one-sample binomial test (α=0.05) to determine if the chatbot's classification accuracy meets the 80% threshold. The analysis of nurse-documented cases will compare the chatbot's risk assessments against gold-standard clinician evaluations. Recommendation quality will be assessed through a dual-method approach. Five clinicians will evaluate 100 chatbot responses using 5-point Likert scales, with scores analyzed via ordinal logistic regression. Concurrently, we will compute automated metrics including BERTScore (semantic similarity), BLEU/METEOR/ROUGE-1 (text generation quality), and Flesch readability scores (targeting FRE $\geq$ 80 and FKGL $\leq$ 6). These metrics will be compared against clinical communication standards.

The comparative assessment of workflow efficiency will employ paired t-tests (or Wilcoxon signed-rank tests for non-normal data) to analyze time savings and satisfaction improvements between traditional and chatbot-based follow-ups. With 40 providers participating (accounting for 20% potential attrition), this sample provides 80% power to detect moderate effect sizes (d=0.5) in the Likert-scale survey responses.

Patient usability evaluation incorporates both quantitative and qualitative methods. Fifty patients will complete standardized surveys assessing ease of use and satisfaction, while in-depth interviews will explore user experiences and preferences. Thematic analysis will identify common patterns and improvement opportunities.

All analyses will report 95% confidence intervals, with p-values < 0.05 considered statistically significant. The mixed-methods approach ensures comprehensive evaluation of both the chatbot's technical performance and its clinical utility in real-world settings.

Ethical considerations

This study will be submitted for approval by the human ethical committees of the Faculty of Dentistry, Chulalongkorn University, as a questionnaire study.

Research schedule

procedure	month											
	1	2	3	4	5	6	7	8	9	10	11	12
Review literature and writing proposal												
Ethics approval												
Data Collection												
Data Processing to Construct a Knowledge Base												
Model Development												
System Integration												
Evaluation												
Research report												
Presentation and writing paper												

Budget

Expense category	Items	Cost (Baht)
Materials	Large Language Model API Access	1,000
	Model training	6,000
	Embedding Model API/Vector Database Subscription	2,000
	Miscellaneous Model Development Costs(Preprocessing tools, debugging, and optimizations)	5,000
Research Report		1,000
Total expenses		15,000

References

1. Thorat V, Rao P, Joshi N, Talreja P, Shetty AR. Role of Artificial Intelligence (AI) in Patient Education and Communication in Dentistry. Cureus. 2024;16(5):e59799.
2. Alhaidry HM, Fatani B, Alrayes JO, Almana AM, Alfhaed NK. ChatGPT in Dentistry: A Comprehensive Review. Cureus. 2023;15(4):e38317.
3. Meenakshi K, Deborah S. An Innovative App (ExoDont) for Postoperative Care of Patients After Tooth Extraction: Prototype Development and Testing study. JMIR PERIOPERATIVE MEDICINE. 2021;4.
4. Zheng X, Zhao J, Wang Z, Jia B. Postoperative online follow-up improves the quality of life of patients who undergo extraction of impacted madibular third molars: a randomized controlled trial. Clinical Oral investigations. 2021.
5. Tuncaroen T. Call Center Scammer Problems. NBTC journal. 2023.
6. Mohammad-Rahimi H, Khoury ZH, Alamdari MI, Rokhshad R, Motie P, Parsa A, et al. Performance of AI chatbots on controversial topics in oral medicine, pathology, and radiology. Oral Surgery Oral Medicine Oral Pathology and Oral Radiology. 2024;137(5):508-14.
7. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 2018;25(9):1248-58.
8. Uribe SE, Maldupa I, Kavadella A, El Tantawi M, Chaurasia A, Fontana M, et al. Artificial intelligence chatbots and large language models in dental education: Worldwide survey of educators. Eur J Dent Educ. 2024;28(4):865-76.
9. Othman Z, Abd Yusof NF, Abdul Malek MH, Zhe Xi MLZX, Rohaizad AN, Amin MF. AI-Powered Dental Consultation Chatbot: Enhancing Public Healthcare Accessibility and Awareness. Journal of Advanced Research in Applied Sciences and Engineering Technology. 2024:14-27.
10. Umer F, Batool I, Naved N. Innovation and application of Large Language Models (LLMs) in dentistry – a scoping review. BDJ Open. 2024;10(1).
11. Goldenthal SB, Portney D, Steppe E, Ghani K, Ellimoottil C. Assessing the feasibility of a chatbot after ureteroscopy. Mhealth. 2019;5:8.
12. Sillmann YM, Monteiro J, Eber P, Baggio AMP, Peacock ZS, Guastaldi FPS. Empowering surgeons: will artificial intelligence change oral and maxillofacial surgery? Int J Oral Maxillofac Surg. 2025;54(2):179-90.
13. Gomez-Cabello CA, Borna S, Pressman SM, Haider SA, Sehgal A, Leibovich BC, et al. Artificial Intelligence in Postoperative Care: Assessing Large Language Models for Patient Recommendations in Plastic Surgery. Healthcare (Basel). 2024;12(11).

14. Bian Y, Xiang Y, Tong B, Feng B, Weng X. Artificial Intelligence-Assisted System in Postoperative Follow-up of Orthopedic Patients: Exploratory Quantitative and Qualitative Study. J Med Internet Res. 2020;22(5):e16896.
15. Nguyen TT, Larrivée N, Lee A, Bilaniuk O. Use of Artificial Intelligence in Dentistry: Current Clinical Trends and Research Advances. J Can Dent Assoc. 2021.
16. Agrawal P, Nikhade P. Artificial Intelligence in Dentistry: Past, Present, and Future. Cureus. 2022;14(7):e27405.
17. Geoghegan L, Scarborough A, Wormald JCR, Harrison CJ, Collins D, Gardiner M, et al. Automated conversational agents for post-intervention follow-up: a systematic review. BJS Open. 2021;5(4).
18. Suarez A, Jimenez J, Llorente de Pedro M, Andreu-Vazquez C, Diaz-Flores Garcia V, Gomez Sanchez M, et al. Beyond the Scalpel: Assessing ChatGPT's potential as an auxiliary intelligent virtual assistant in oral surgery. Comput Struct Biotechnol J. 2024;24:46-52.
19. Jeong H, Han SS, Yu Y, Kim S, Jeon KJ. How well do large language model-based chatbots perform in oral and maxillofacial radiology? Dentomaxillofac Radiol. 2024;53(6):390-5.
20. Dwyer T, Hoit G, Burns D, Higgins J, Chang J, Whelan D, et al. Use of an Artificial Intelligence Conversational Agent (Chatbot) for Hip Arthroscopy Patients Following Surgery. Arthrosc Sports Med Rehabil. 2023;5(2):e495-e505.
21. Sousa CS, Turrini RNT. Complications in orthognathic surgery: A comprehensive review. Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology. 2012;24(2):67-74.
22. Arian MSH, Sifat FA, Ahmed S, Mohammed N, Farook TH, Dudley J. Dental Loop Chatbot: A Prototype Large Language Model Framework for Dentistry. Software. 2024;3(4):587-94.
23. Steybe D, Poxleitner P, Aljohani S, Herlofson BB, Nicolatou-Galitis O, Patel V, et al. Evaluation of a context-aware chatbot using retrieval-augmented generation for answering clinical questions on medication-related osteonecrosis of the jaw. J Craniomaxillofac Surg. 2025.
24. Batool I, Naved N, Kazmi SMR, Umer F. Leveraging Large Language Models in the delivery of post-operative dental care: a comparison between an embedded GPT model and ChatGPT. BDJ Open. 2024;10(1):48.
25. Semerci ZM, Yardimci S. Empowering Modern Dentistry: The Impact of Artificial Intelligence on Patient Care and Clinical Decision Making. Diagnostics (Basel). 2024;14(12).
26. Miragall MF, Knoedler S, Kauke-Navarro M, Saadoun R, Grabenhorst A, Grill FD, et al. Face the Future-Artificial Intelligence in Oral and Maxillofacial Surgery. J Clin Med. 2023;12(21).
27. Panahi O. AI: A New Frontier in Oral and Maxillofacial Surgery. Acta Scientific Dental Sciences 2024;8.
28. Aminoshariae A, Kulild J, Nagendrababu V. Artificial Intelligence in Endodontics: Current Applications and Future Directions. J Endod. 2021;47(9):1352-7.
29. Abdalla AQ, Aziz TA. ChatGPT: A game-changer in oral and maxillofacial surgery. Journal of Medicine, Surgery, and Public Health. 2024;2.