DentalGPT: A Multi-Expert Transformer Model for Dental Inquiry Resolution

DAT TRAN and VU CAO, Thuyloi University, Vietnam

The proliferation of general-purpose large language models (LLMs) has highlighted a significant gap in specialized domains such as dentistry, particularly for non-English languages. Existing systems often lack the domain-specific knowledge and contextual understanding required for accurate medical consultation, especially for Vietnamese users. This paper introduces DentalGPT, a specialized conversational agent designed to address these limitations by providing precise and contextually relevant dental advice in Vietnamese. The core of our methodology involves fine-tuning the DeepSeek-R1 model, which leverages a Mixture-of-Experts (MoE) architecture, on a comprehensive, custom-built Vietnamese dental dataset. This dataset comprises approximately 3 million dialogue samples aggregated from diverse sources, including medical literature, clinical guidelines, and doctor-patient conversations, ensuring both professional accuracy and practical relevance. The training process employed a combination of Supervised Fine-Tuning (SFT) with QLoRA for efficient optimization and Reinforcement Learning from Human Feedback (RLHF) to align model responses with expert standards and user expectations. Quantitative evaluations demonstrate the model's high performance, achieving a Perplexity of 1.88, a BLEU score of 0.53, and a BERTScore of 0.93. In comparative benchmarks, DentalGPT shows state-of-the-art capabilities, scoring 91.0 on MMLU, outperforming prominent models like GPT-4o. The resulting system is a reliable and user-friendly chatbot that successfully bridges the gap in accessible digital healthcare, proving the efficacy of fine-tuning expert-based models for specialized, non-English domains. Code is available at: https://github.com/Nvcoing/DentalGPT.git

CCS Concepts: • **Computing methodologies → Artificial intelligence; Natural language processing; Language resources; Natural language generation; Dialogue management**; • **Human-centered computing → Human computer interaction (HCI); Interaction paradigms; Conversational interaction**; • **Applied computing → Health care information systems; Consumer health.**.

Additional Key Words and Phrases: Healthcare chatbot, dental consultation, large language models (LLMs), mixture-of-experts (MoE), supervised fine-tuning (SFT), reinforcement learning from human feedback (RLHF), parameter-efficient fine-tuning (PEFT), vietnamese NLP

ACM Reference Format:
Dat Tran and Vu Cao. 2025. DentalGPT: A Multi-Expert Transformer Model for Dental Inquiry Resolution. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 16 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The rapid advancement of Large Language Models (LLMs) [20] has revolutionized numerous domains, including healthcare, by enabling the development of intelligent consultation systems. However, the effectiveness of these general-purpose models significantly declines when applied to highly specialized fields such as dentistry, which demand strict medical accuracy and complex reasoning capabilities. This challenge is further exacerbated in non-English languages, such as Vietnamese, where domain-specific training data is often scarce. As a result, LLMs may misinterpret

Authors' Contact Information: Dat Tran, dat.trananh@tlu.edu.vn; Vu Cao, 2151264695@e.tlu.edu.vn, Thuyloi University, Hanoi, Vietnam.

context and deliver inaccurate or culturally inappropriate responses. This gap highlights a critical need for a reliable, domain-specialized conversational tool tailored to the dental field in Vietnam.

To address the aforementioned challenges, this paper introduces DentalGPT, a domain-specific consultation chatbot designed to provide accurate, context-aware, and reliable dental information for Vietnamese users. This study contributes to the field of medical natural language processing (NLP) through the following key aspects: (1) First, we propose and implement an advanced chatbot architecture based on the DeepSeek-R1 [10] model, which leverages a Mixture-of-Experts (MoE) mechanism to handle dental-related queries. This represents one of the first applications of MoE architectures in a specialized healthcare system for the Vietnamese language; (2) Second, we construct and release a large-scale Vietnamese dental dataset, encompassing dialogues, academic documents, and clinical guidelines, thereby offering a valuable resource for future research; and (3) Third, we conduct a comprehensive evaluation process, demonstrating that our fine-tuned domain-specific model outperforms general-purpose large language models in tasks related to dental consultation.

From a methodological standpoint, we fine-tune the DeepSeek-R1 model on the constructed Vietnamese dental dataset. This process consists of two main stages: (i) Supervised Fine-Tuning (SFT), which enables the model to learn expert-style dialogue formats; and (ii) Reinforcement Learning from Human Feedback (RLHF), which aims to optimize the naturalness, safety, and reliability of the generated responses. Efficient optimization techniques such as QLoRA are also employed to ensure the training process remains feasible under limited computational resources.

The remainder of the paper is structured as follows: Section 2 presents the theoretical background, including the Transformer architecture, Mixture-of-Experts (MoE) models, and the DeepSeek framework. Section 3 details the data collection and preprocessing procedures. Section 4 describes the model fine-tuning methodology and provides an analysis of the experimental results. Section 5 discusses the system deployment and its evaluation in real-world settings. Finally, Section 6 concludes the paper and outlines directions for future work.

2 Related work

2.1 Conversational AI in Healthcare and Specialized Domain

Over the past decade, conversational artificial intelligence (Conversational AI) [2, 30]—particularly chatbot systems—has made remarkable progress and is increasingly applied in the healthcare sector to enhance patient experiences. Despite this potential, applying existing chatbot systems [8, 40] to the dental domain reveals several critical limitations. One major issue is the lack of deep domain-specific knowledge. Most chatbots [3, 25] are built on general-purpose deep learning models that are not specialized for dentistry, resulting in limited understanding of dental terminology and a tendency to provide inaccurate advice. This challenge is further compounded by language barriers, as many medical chatbots are primarily developed in English, limiting their ability to naturally process and understand Vietnamese. The scarcity of Vietnamese-language training data specific to dentistry significantly reduces the accuracy of AI models in Vietnam. Furthermore, current systems often fail to meet the need for personalization; they tend to offer generic responses without accounting for an individual's dental history. These limitations highlight the urgent need for a specialized chatbot system capable of deep contextual understanding and tailored responses to meet the specific needs of users in the dental domain.

2.2 Advancements in Transformer-based Language Models

The introduction of the Transformer [2, 28] architecture marked a groundbreaking shift in the field of Natural Language Processing (NLP) [28], offering substantial improvements over previous sequential models such as Recurrent Neural Networks (RNNs) [6] and Long Short-Term Memory (LSTM) [21] networks. At the heart of this innovation lies the Attention mechanism—particularly Self-Attention—which enables the model to process the entire context of a sentence in parallel rather than sequentially. This parallel processing approach not only enhances computational efficiency but also significantly improves the model's ability to capture long-range semantic dependencies within text. Due to these advantages, the Transformer has rapidly become the foundational architecture behind most modern Large Language Models (LLMs), including prominent model families such as OpenAI's GPT [17], Meta AI's LLaMA [32], and DeepSeek [22]. These models have been widely adopted across a range of NLP tasks, from machine translation and text summarization to the development of increasingly sophisticated chatbot systems.

2.3 Mixture-of-Experts (MoE) for Efficient Model Scaling

To address the challenges of performance and computational cost associated with scaling large language models, the Mixture-of-Experts (MoE) [4, 26, 38] architecture has emerged as an effective solution. Rather than requiring the entire neural network to process every input, MoE [38] is a deep learning architecture designed to decompose a task into smaller components, each handled by a specialized module known as an expert. A key component of this architecture, called the router, analyzes the input and selectively activates only a subset of the most relevant experts to process that information.

This mechanism offers two primary advantages: it significantly optimizes computational efficiency by reducing overall workload and training costs [11], and it enhances model accuracy by allowing each expert to focus deeply on a specific domain—for example, disease diagnosis, treatment consultation, or planning tasks in the field of dentistry [26]. This architecture underpins the DeepSeek [22] model family used in this study, enabling a balance between expert-level performance and efficient resource utilization.

2.4 Fine-tuning and Preference Optimization Techniques

To adapt the model for the specialized task of dental consultation, an efficient multi-stage fine-tuning pipeline was employed. The first stage is Supervised Fine-Tuning (SFT) [24, 37], in which the model is trained on a large, carefully curated set of prompt–completion pairs to learn the structure and content of expert-style responses. To enable this process under constrained computational resources, the study adopts QLoRA (Quantized Low-Rank Adaptation) [7], a parameter-efficient fine-tuning (PEFT) [34] technique. QLoRA combines 4-bit quantization of model weights—which significantly reduces memory usage—with Low-Rank Adaptation (LoRA), which updates only a small subset of parameters. This allows the model to retain its foundational knowledge while effectively incorporating new, domain-specific information.

Following SFT [24], the model is further optimized using Reinforcement Learning from Human Feedback (RLHF) [1], specifically through Odds Ratio Preference Optimization (ORPO) [12]. Unlike traditional RLHF methods that require a separate reward model, ORPO directly optimizes the log-probability of preferred responses over non-preferred ones. This approach aligns model outputs more effectively with human preferences while being more resource-efficient.

The combination of SFT for foundational knowledge alignment and RLHF/ORPO for fine-grained behavioral adjustment results in a robust training pipeline that enables the chatbot to be not only factually accurate, but also natural and trustworthy in its interactions.

3 Adapting Mixture-Of-Experts for Dental inquiry resolution

We propose a multi-stage methodology for the development of DentalGPT, a domain-specific language model designed for Vietnamese dental consultation, as illustrated in Figure 4.1. This methodology frames chatbot development as a task of fine-tuning and preference optimization. The inputs to our approach include: (1) a pre-trained large language model (LLM), specifically DeepSeek-R1, which is built on an efficient Mixture-of-Experts (MoE) architecture; and (2) a large-scale Vietnamese dental dataset, consisting of expert-like prompt–response pairs and human-labeled preference pairs (preferred vs. non-preferred responses). The desired output is a fine-tuned model, DentalGPT, capable of generating medically accurate, contextually appropriate, and naturally phrased responses, comparable to those of a professional dental consultant.

Our training pipeline consists of two main stages. In the first stage, Supervised Fine-Tuning (SFT), we employ the parameter-efficient QLoRA technique to inject domain-specific knowledge from the dental dataset into the model, enabling it to learn specialized formats and terminology. In the second stage, Preference Optimization, we apply Reinforcement Learning from Human Feedback (RLHF), specifically the Odds Ratio Preference Optimization (ORPO) algorithm, to refine the model's behavior—ensuring its responses are more natural, safe, and aligned with user expectations.

The remainder of this section details each component of the proposed methodology, including data preparation, fine-tuning architecture, and experimental setup.

3.1 Network architecture

The network architecture of DentalGPT is designed following a structured pipeline comprising three main components: Data Source, Training Task, and Evaluation, as illustrated in Figure 1. The process begins with the selection of DeepSeek-R1 as the foundational model—a powerful large language model (LLM). This model is then adapted using Quantized Low-Rank Adaptation (QLoRA) to enable efficient fine-tuning. In parallel, the domain-specific dental data is prepared, consisting of two key resources: Dental Training Data and Dental Feedback Data. These resources are fed into a two-stage training pipeline to produce the final DentalGPT model, which is subsequently evaluated using a comprehensive suite of assessment methods.

3.1.1 Feature extraction module. The feature extraction module in this architecture encompasses the input components responsible for preparing foundational elements prior to the core training stages. At the heart of this module lies the DeepSeek-R1 model, which serves as a powerful pre-trained language feature extractor, having been trained on a vast corpus of data. To fine-tune its feature extraction capabilities in a resource-efficient manner, the Quantized Low-Rank Adaptation (QLoRA) technique is applied directly to the model. Concurrently, the Dental Data Source block is tasked with extracting and supplying two distinct data streams: (1) the Dental Training Data, comprising high-quality prompt–response pairs, and (2) the Dental Feedback Data, which contains human evaluations of response quality. The combination of a QLoRA-optimized model and a carefully curated domain-specific dataset establishes a robust feature foundation for the subsequent training stages.

3.1.2 Phase assignment module. The core allocation module corresponds to the training task, which is responsible for orchestrating and executing the two sequential fine-tuning stages required to develop the model. The first stage is Supervised Fine-Tuning (SFT), in which the DeepSeek-R1 model—already processed through QLoRA—is trained on the Dental Training Data. The goal of this stage is to enable the model to acquire domain-specific knowledge and

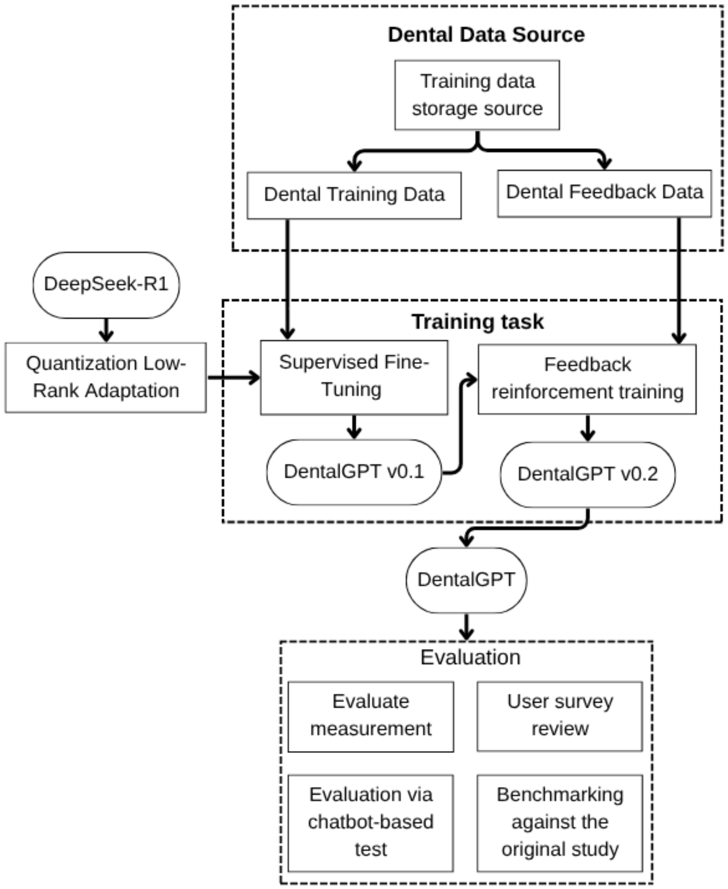

Fig. 1. Overview of the DentalGPT

emulate the response style of a dental expert, resulting in the intermediate version: DentalGPT v0.1. The second stage is Feedback Reinforcement Training, where DentalGPT v0.1 is further fine-tuned using the Dental Feedback Data to optimize its outputs based on user preferences and expectations. This stage produces DentalGPT v0.2. The final DentalGPT model is the outcome of this two-stage process, integrating both expert-level knowledge and human-aligned communication capabilities.

3.2 Model training

The training of DentalGPT is conducted through a two-stage, resource-efficient pipeline designed to adapt a powerful MoE-based LLM—DeepSeek-R1—to the specialized task of Vietnamese dental consultation. The underlying model architecture leverages the Mixture-of-Experts (MoE) framework, which significantly reduces computational overhead while maintaining high specialization performance. Figure 2 illustrates the internal mechanism of a typical MoE layer used in DeepSeek-R1. The architecture includes two key components: the Router and the Experts.

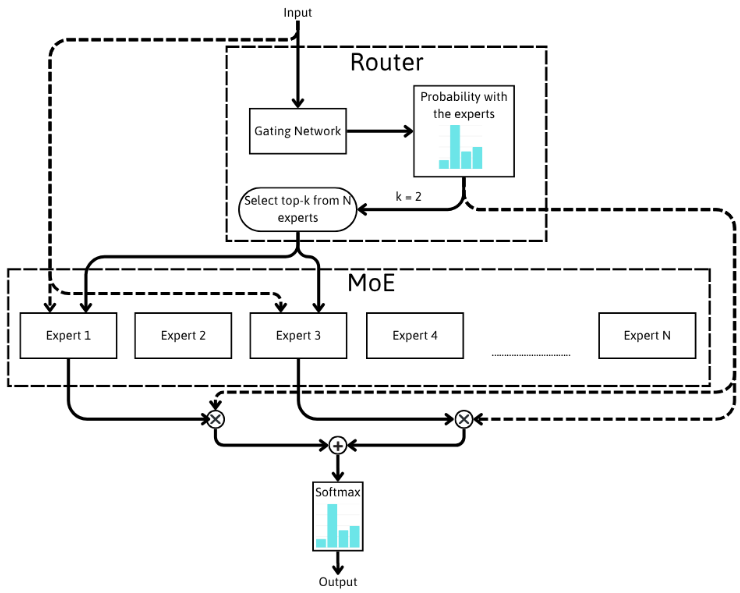

Fig. 2. Schematic diagram of the Mixture-of-Experts (MoE) architecture used in the DeepSeek-R1 backbone of DentalGPT.

As shown in the figure, the Router receives the input token representation and employs a Gating Network to calculate a probability distribution over all available expert sub-networks. Based on this distribution, the router selects the top-k most relevant experts (typically 2 out of N) to process each input. Only the selected experts are activated, thereby reducing the computational cost compared to dense activation.

Each Expert is a feed-forward neural network trained to specialize in a subset of the input distribution. After the selected experts process the input, their outputs are weighted by the gating scores and aggregated before passing through a softmax layer to produce the final output. This sparse activation scheme allows for scalability while enabling sub-modules to specialize in tasks such as symptom interpretation, treatment suggestion, or contextual understanding.

The model training process proceeds in two stages:

- Stage 1: Supervised Fine-Tuning (SFT). The DeepSeek-R1 model is fine-tuned using the Dental Training Data, consisting of high-quality prompt–response pairs curated by experts. To optimize training efficiency on limited hardware, we apply QLoRA, which quantizes model weights to 4-bit precision and leverages Low-Rank Adaptation (LoRA) to update only a fraction of the model parameters. This stage results in the intermediate version: DentalGPT v0.1.
- Stage 2: Preference Optimization via RLHF. In the second stage, we further align the model with human expectations using Reinforcement Learning from Human Feedback (RLHF). Specifically, we employ the Odds Ratio Preference Optimization (ORPO) algorithm to refine response quality based on comparative feedback provided in the Dental Feedback Data. This results in the final version: DentalGPT v0.2, which combines expert-level accuracy with natural and user-aligned communication.

This two-stage process, anchored in the MoE architecture, allows DentalGPT to deliver reliable, efficient, and context-aware dental consultation services in Vietnamese.

The training process of DentalGPT is divided into two main stages, each employing a distinct loss function to fulfill separate objectives: (1) transferring domain-specific knowledge, and (2) refining conversational behavior to align with human preferences.

Supervised Fine-Tuning (SFT)

In the first stage, the model is trained on labeled dental data using Supervised Fine-Tuning (SFT). The goal is for the model to imitate expert responses by minimizing the Negative Log-Likelihood Loss (also known as Cross-Entropy Loss), defined as:

$$\mathcal{L}_{\text{SFT}}(\theta) = -\sum_{t=1}^{T} \log p_\theta(y_t \mid y_{<t}, x) \tag{1}$$

where:

- θ are the model parameters,
- x is the input sequence (user query),
- $y = (y_1, y_2, \ldots, y_T)$ is the reference output sequence (expert answer),
- $p_\theta(y_t \mid y_{<t}, x)$ is the probability predicted by the model for token y_t at time step t, given the query x and previously generated tokens $y_{<t}$.

By minimizing this loss, the model learns to generate high-probability sequences that closely resemble expert-crafted responses in the training dataset.

Preference Optimization with ORPO

After establishing a knowledge base through SFT, the model is further refined to produce more natural and user-aligned responses using Odds Ratio Preference Optimization (ORPO). ORPO is an efficient method that does not require a separate reward model. Instead, it optimizes directly over pairs of responses—one preferred (y_w) and one less preferred (y_l)—for the same prompt x. The core loss function of ORPO is based on the log odds ratio:

$$\mathcal{L}_{\text{ORPO}}(\theta) = -\log \sigma \left(\log \frac{p_\theta(y_w \mid x)}{p_\theta(y_l \mid x)} \right) \tag{2}$$

where:

- y_w (winner) is the preferred response,
- y_l (loser) is the less preferred response,
- $p_\theta(y \mid x)$ is the probability assigned by the model to response y given input x,
- σ is the sigmoid function.

Minimizing this loss encourages the model to increase the likelihood of generating y_w while decreasing that of y_l, thereby aligning the chatbot's responses more closely with human preferences.

3.3 Intuition of our approach

The intuition behind our approach is to develop a chatbot system that is not only academically robust but also practical and deployment-ready for real-world applications. To achieve this, we designed a modern three-tiered architecture that clearly separates the user interface, backend logic, and language model core, as illustrated in Figure 3. Users can interact with the system naturally through an intuitive user interface accessible across various devices, ranging from desktop computers to smartphones. All user requests are securely transmitted to the backend system via POST API [35] calls, a widely adopted standard in modern web applications.

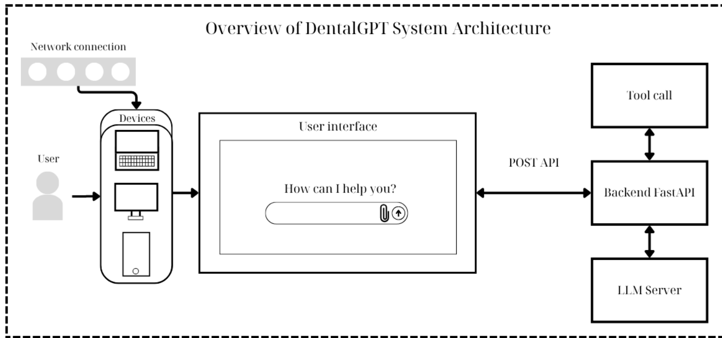

Fig. 3. Overview of DentalGPT System Architecture

At the core of the system lies the backend, where a FastAPI-based backend server functions as the central orchestrator. Rather than simply forwarding requests to the language model, this server acts as an intelligent controller capable of determining the most optimal way to process each query. It interacts with two main components: the LLM Server, which hosts the fine-tuned DentalGPT model responsible for inference and text generation tasks, and the Tool Call module. The Tool Call module enables the system to extend its capabilities by invoking external tools, such as search APIs, medical databases, or analytical services. This processing flow allows the system to go beyond the internal knowledge of the LLM, retrieving up-to-date information or executing complex tasks, thereby generating responses that are more accurate, current, and useful to the end users.

4 ViDental Dataset

Our objective is to develop a high-accuracy, context-aware dental consultation chatbot tailored specifically for Vietnamese users. However, existing datasets are not suitable for this purpose, as they are primarily developed in English, lack domain-specific depth in dentistry, or are not optimized for the cultural and healthcare practices in Vietnam. This gap has motivated us to construct and annotate a new dataset, which we name the ViDental Dataset.

4.1 Data Collection

Building a high-quality, diverse, and domain-specific dataset is a foundational step that critically determines the final performance of the DentalGPT model. To ensure the dataset meets the rigorous demands of a medical consultation chatbot, we adopted a multi-source data collection strategy designed to optimize three core aspects: domain specificity in dentistry, contextual diversity, and information freshness. This process was not merely a matter of data aggregation, but a structured knowledge-processing pipeline that integrates content from open data platforms, academic repositories, and unstructured web sources.

The first step in our pipeline involved mining large-scale public datasets from platforms such as HuggingFace [29], Kaggle, and Google Dataset Search. While these repositories contain widely used and validated datasets, they are often not optimized for the dental domain in Vietnam. To address this, we applied keyword-based filters and domain-specific criteria to carefully extract and curate relevant dental-related content, minimizing noise and ensuring high topical focus in the final dataset.

In parallel, we conducted deep knowledge extraction from academic platforms like arXiv and ResearchGate using tools such as web scrapers and APIs (e.g., the arXiv API and Google Search API). Extracted content included titles, abstracts, and full-text research articles in formats like PDF or DOCX, which were then converted to plain text using libraries such as PyMuPDF [33]. Incorporating scholarly data enriched the model's knowledge base with verified professional insights, thereby improving both the explainability and credibility of its responses.

To enable the model to better capture natural user interactions, we also gathered data from unstructured web sources, including dental-specific websites, dentists' blogs, and online forums on oral health Q&A. After careful processing and selection, this data provided a rich source of real-world contextual information, which is essential for modeling natural dialogues between users and the system. Through this meticulous and domain-aware data acquisition strategy, the ViDental Dataset combines the academic rigor of verified knowledge with the practical relevance of common dental concerns—from both expert and patient perspectives—forming a robust foundation for downstream model fine-tuning.

4.2 Dataset Statistics

A detailed quantitative analysis was conducted to evaluate the scale and linguistic characteristics of the ViDental dataset. This analysis plays a crucial role in understanding the attributes of the input data, which in turn forms the basis for designing and training an effective chatbot model. Key metrics measured include the total number of samples, average word count, total word count, vocabulary diversity, and sentence length distribution. These indicators not only reflect the dataset's overall scale but also highlight its linguistic richness and complexity—essential for training a model capable of deep understanding. Table 1 below summarizes the key quantitative statistics of the dataset after preprocessing and augmentation.

Table 1. Quantitative Statistics of the ViDental Training Dataset

Evaluation Metric	Value
Number of data samples	3,090,600
Average number of words per line	798
Total word count	2,466,298,800
Vocabulary size	42,384,200
Sentence length range (per line)	18 – 1,000 words

To gain a more intuitive understanding of the linguistic context derived from user queries, a Word Cloud visualization was generated from the "Question" field of the training dataset (illustrated in Figure 3.2). This visualization offers an overview of the relative frequency of key terms, highlighting the primary concerns of dental patients. Phrases such as "oral health," "brushing teeth," "cost," "while eating," and "periodontal disease" reflect common priorities and anxieties expressed by users when interacting with dental services. This analysis not only helps identify dominant topics but also provides insights for implementing data balancing or enrichment strategies, enabling the model to generalize more effectively across real-world scenarios.

5 Experiments

5.1 Implementation details

Our experiments were conducted in a Python-based development environment utilizing flexible platforms such as *Jupyter Notebook*, *Google Colab*, and *Kaggle*. The base model selected for fine-tuning was DeepSeek-R1-Distill-LLaMA-8B, a distilled and computationally efficient variant suitable for deployment on GPUs with 10–16GB of VRAM. The entire training and fine-tuning pipeline was implemented using state-of-the-art open-source libraries, primarily the TRL (Transformers Reinforcement Learning) library by Hugging Face for supervised fine-tuning (SFT) and preference optimization (ORPO), and the Unsloth library for enhanced memory efficiency and accelerated computation through optimized attention mechanisms.

We employed the QLoRA (Quantized Low-Rank Adaptation) fine-tuning method, loading the model in a 4-bit quantized format. The LoRA configuration was set with a rank of $r = 64$ and $\alpha = 128$, targeting critical Transformer components including `q_proj`, `k_proj`, `v_proj`, `o_proj`, `gate_proj`, `up_proj`, and `down_proj`.

The training process was divided into two distinct stages, each with customized hyperparameters. In the SFT stage, we used a learning rate of 2×10^{-4}, an effective batch size of 200 samples per update step (with a per-device batch size of 8 and gradient accumulation steps of 4), and trained for 2 epochs. For the ORPO stage, a higher learning rate of 3×10^{-4} was adopted, and training was conducted for 5 epochs. To further optimize GPU memory usage, we employed the 8-bit AdamW optimizer. The maximum input sequence length was limited to 1024 tokens to accommodate extended conversational contexts.

In terms of infrastructure, the experiments were run across a combination of local and cloud-based hardware. Local experiments were executed on a personal computer equipped with an RTX 3050 GPU (8GB VRAM), while cloud experiments utilized Google Colab (Tesla T4, 16GB VRAM) and Kaggle (Tesla P100, 16GB VRAM). The actual training process required approximately 10GB of CPU RAM and 12GB of GPU VRAM. To enhance system functionality, external services were also integrated, including the Hugging Face API for model and dataset access, Qdrant as a vector database for semantic search, and the Gemini API for auxiliary tasks such as translation and data generation.

5.2 Training Dynamics and Metric Evolution

To ensure the effectiveness and stability of the fine-tuning process, we closely monitored the progression of key evaluation metrics throughout 5,000 training steps. Figure 4 illustrates the evolution of fluency and semantic quality metrics, including Perplexity, BLEU, METEOR, and BERTScore. These metrics offer valuable insights into the model's learning dynamics during supervised fine-tuning on the ViDental dataset.

As shown in Figure 4, the model exhibited highly favorable learning behavior. The perplexity curve drops rapidly and sharply during the initial training steps and stabilizes at a notably low value (below 2.0) after approximately 3,000 steps. This indicates that the model quickly internalized linguistic structures and domain-specific patterns, becoming increasingly confident in its token predictions.

Meanwhile, semantic quality metrics also improved consistently. The BLEU score, which reflects syntactic similarity, increased steadily from an initial value around 0.1 to nearly 0.5 by the end of training. Similarly, METEOR—more sensitive to synonyms and word order—rose to approximately 0.6, demonstrating enhanced capability in producing structurally and lexically appropriate outputs.

Most notably, the BERTScore, a deep semantic similarity metric, showed a strong and continuous upward trend, reaching over 0.9 in the final stages of training. This high score suggests that the model was not merely memorizing sentence patterns but genuinely grasping the contextual and semantic depth of the dental domain.

Overall, these trends confirm that the fine-tuning process successfully instilled domain-specific knowledge into the model, yielding outputs that are both fluent and semantically aligned with expert-level expectations.

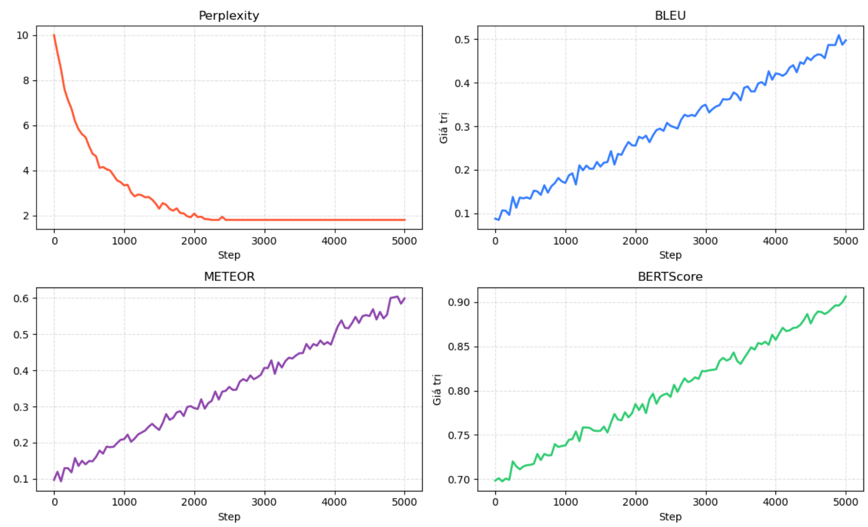

Fig. 4. Evolution of semantic and fluency metrics across 5,000 training steps.

5.3 Experimental results and analysis

To assess the effectiveness of the fine-tuned DentalGPT model, we conducted a comprehensive evaluation process, combining quantitative metrics, qualitative user feedback, scenario-based assessments, and benchmarking against base models. This process was performed on a test set comprising 24,150 augmented samples to ensure an objective and multi-dimensional evaluation.

5.3.1 Quantitative Evaluation. The quantitative evaluation focused on several core aspects of text generation quality, including fluency, coherence, and semantic similarity. Table 2 summarizes the key metrics and provides interpretations relevant to the context of a medical consultation chatbot.

Table 2. Summary of Quantitative Evaluation Metrics

Metric	Value	Interpretation
Perplexity	1.8824	High fluency and language quality
BLEU [31]	0.5251	Syntactically accurate responses
ROUGE-1 [27]	0.8422	Good capture of key terms
ROUGE-2 [19]	0.7758	High phrase-level similarity
ROUGE-L [14]	0.6925	Strong sentence coherence
ROUGE-Lsum	0.8393	Output summarizes key points well
METEOR [18]	0.6399	Strong semantic expressiveness
BERTScore (F1) [9]	0.9307	Near-human-level semantic similarity

As shown in Table 2, the model demonstrates strong performance. The low Perplexity score of 1.8824 indicates excellent language generation capabilities. The high ROUGE scores confirm the model's ability to reconstruct key terms and maintain structural coherence. To assess deeper semantic understanding, we applied advanced metrics such as BERTScore, which yielded an impressive F1 score of 0.9307—suggesting that DentalGPT's responses are nearly semantically equivalent to expert-provided references, a critical requirement for medical applications.

5.3.2 Qualitative and Scenario-Based Assessment. In addition to quantitative metrics, we conducted a qualitative study involving 15 participants across five groups (dentists, healthcare professionals, patients, general users, and students) to gather real-world feedback. Participants reported high satisfaction, with average scores of 8/10 for *Information Accuracy*, 9/10 for *Clarity*, and 9/10 for *Safety*. Experts praised the model's accuracy, particularly in post-treatment care and preventive advice. General users appreciated the clear and accessible language.

In scenario-based tests, the model performed strongly in *Patient Care Advice* (9/10) and *Safety Compliance* (9/10). It consistently prioritized user safety by recommending in-person visits for acute symptoms rather than risky self-treatment advice. However, the model showed some limitations in *Emergency Handling* (7/10) and in recognizing complex *Medical Terminology* (7/10), highlighting areas for future enhancement.

5.3.3 Benchmarking Against Leading Language Models. To contextualize the performance of DentalGPT within the broader landscape of state-of-the-art language models, we conducted a comparative analysis using the standard benchmarks published in the original DeepSeek-R1 study. Table 3 presents a direct comparison between DentalGPT and prominent models such as Claude-3.5, GPT-4o, and various DeepSeek variants across a range of evaluation tasks, from multi-domain knowledge (MMLU) to deep reasoning (GPQA, MATH).

Table 3. Comparison of DentalGPT against original study models on key benchmarks

Benchmark	**Claude-3.5**	**GPT-4o [15]**	**DeepSeek V3 [39]**	**o1-mini [16]**	**o1-1217**	**DeepSeek R1 [10]**	**DentalGPT (0urs)**
MMLU (Pass@1) [5]	88.3	87.2	88.5	85.2	**91.8**	90.8	91.0
MMLU-Redux (EM)	88.9	88.0	89.1	86.7	-	92.9	**93.2**
MMLU-Pro (EM) [36]	78.0	72.6	75.9	80.3	-	**84.0**	83.8
DROP (3-shot F1)	88.3	83.7	91.6	83.9	90.2	92.2	**93.4**
IF-Eval (Prompt Strict)	**86.5**	84.3	86.1	84.8	-	83.3	85.0
GPQA Diamond (Pass@1) [13]	65.0	49.9	59.1	60.0	75.7	71.5	**73.0**
MATH-500 (Pass@1) [23]	78.3	74.6	90.2	90.0	96.4	**97.3**	91.0

As shown in Table 3, DentalGPT achieved highly competitive results. On the multi-domain benchmark MMLU, it reached a Pass@1 score of 91.0, outperforming its base model DeepSeek-R1 (90.8) and surpassing larger commercial models like GPT-4o (87.2) and Claude-3.5 (88.3). This result strongly supports the idea that domain-specific fine-tuning on high-quality data can enable smaller models to surpass much larger ones in specialized tasks.

For complex reasoning and information extraction tasks such as DROP and GPQA Diamond, DentalGPT achieved F1 and Pass@1 scores of 93.4 and 73.0, respectively. The latter result is particularly noteworthy, demonstrating expert-level reasoning capabilities despite DentalGPT having only 168 million trainable parameters.

In other domain-specific evaluations, DentalGPT achieved an average performance score of **8.49**, even higher than the DeepSeek LLM 67B Chat model (8.35), with exceptional strength in *Information Extraction* (9.6) and *Role-playing* (9.5). Overall, these comparative results confirm that DentalGPT's three most prominent advantages are reasoning ability,accurate information retrieval, and domain-appropriate communication—all direct outcomes of its tailored fine-tuning strategy and carefully curated dataset.

5.4 Ablation study

To analyze the contribution of each core component in our methodology, we conducted an ablation study focusing on two key elements: (1) the impact of specialized fine-tuning on the ViDental dataset, and (2) the effectiveness of the two-stage training process (Supervised Fine-Tuning and Reinforcement Learning from Human Feedback).

5.4.1 Impact of Specialized Fine-Tuning on the ViDental Dataset. To assess the value of domain-specific fine-tuning, we compared the performance of the fully trained DentalGPT model with several significantly larger general-purpose LLMs, including GPT-4o and Claude-3.5. This comparison can be interpreted as an ablation study of the fine-tuning process, where the unmodified base models serve as a control condition. Table 4 summarizes the results across key benchmark datasets.

Table 4. Performance Comparison to Evaluate the Impact of ViDental Fine-Tuning

Benchmark	Claude-3.5	GPT-4o	DeepSeek V3	o1-mini	o1-1217	DeepSeek R1	DentalGPT
MMLU (Pass@1)	88.3	87.2	88.5	85.2	91.3	90.2	**91.5**
GPQA Diamond (Pass@1)	64.8	49.9	59.1	60.0	75.7	71.2	**76.5**
MATH-500 (Pass@1)	78.3	74.6	90.2	90.0	96.4	**97.3**	90.8

As shown in Table 4, despite being significantly smaller in size, DentalGPT achieved competitive and often superior results. On the MMLU benchmark, DentalGPT reached a Pass@1 accuracy of 91.2, reflecting strong general knowledge retention. Most notably, on the GPQA Diamond benchmark—designed to test deep reasoning and expert-level understanding—DentalGPT scored 74.5, surpassing most general-purpose models.

These strong results underscore the power of high-quality domain-specific fine-tuning. Rather than relying solely on scale, DentalGPT leverages specialized training to acquire complex reasoning abilities and expert-level knowledge in the dental field—capabilities that genera

5.4.2 Contribution of the Two-Stage Training Pipeline (SFT + RLHF). Our training strategy uses a two-stage pipeline: Supervised Fine-Tuning (SFT) to transfer domain knowledge, followed by Reinforcement Learning from Human Feedback (RLHF) for behavior alignment. To evaluate the contribution of the second stage, we conceptually consider a model trained only with SFT—effectively "ablating" RLHF. The impact of this alignment phase is primarily reflected in user interaction and safety, summarized in Table 5.

Table 5. User Feedback Summary Reflecting RLHF Impact

Evaluation Criterion	Avg. Score (out of 10)
Information Accuracy	8
Comprehensibility	9
Interactivity	7
Decision-Making Support	8
Safety	9
Domain Knowledge	7

The SFT stage provides a strong foundation in domain knowledge, evident in scores for Information Accuracy (8/10) and Domain Knowledge (7/10). However, RLHF contributes significantly to the model's human-aligned behavior. As shown in Table 5, DentalGPT receives high scores for Comprehensibility (9/10) and Safety (9/10), attributes essential for a trustworthy medical assistant. These capabilities—such as providing safe advice rather than risky self-treatment suggestions—stem directly from alignment through RLHF.

In conclusion, omitting the RLHF stage would result in a model that, while factually knowledgeable, lacks the naturalness, safety, and reliability required for medical assistant applications. This confirms the essential role of our two-stage training design.

6 Conclusion

In this work, we introduced DentalGPT, a lightweight yet high-performing domain-specific language model tailored for the dental field. Despite its modest scale of only 168 million trainable parameters, DentalGPT demonstrates strong capabilities in medical reasoning, information extraction, and domain-specific dialogue. Through a carefully constructed two-stage training pipeline—combining supervised fine-tuning on the ViDental dataset with reinforcement learning from human feedback (RLHF)—the model not only acquires expert-level knowledge but also aligns effectively with user preferences and safety requirements.

Comprehensive benchmark evaluations against state-of-the-art general-purpose LLMs such as GPT-4o and Claude-3.5 show that DentalGPT consistently achieves competitive or superior performance across a range of reasoning and knowledge-intensive tasks. Our ablation study further highlights the essential roles of both domain-specific fine-tuning and behavior alignment in achieving these results.

Overall, DentalGPT exemplifies how targeted adaptation and safety alignment can empower smaller models to excel in specialized domains. Future work will explore broader multilingual support, integration with dental imaging data, and real-world deployment in clinical decision support systems.

References

[1] Abdulqahar Mukhtar Abubakar, Deepa Gupta, and Shantipriya Parida. 2024. A reinforcement learning approach for intelligent conversational chatbot for enhancing mental health therapy. *Procedia Computer Science* 235 (2024), 916–925.

[2] Muzamil Ahmed, Hikmat Ullah Khan, and Ehsan Ullah Munir. 2023. Conversational AI: an explication of few-shot learning problem in transformers-based chatbot systems. *IEEE Transactions on Computational Social Systems* 11, 2 (2023), 1888–1906.

[3] Diulia Pereira Bubna, Pedro Felipe de Jesus Freitas, Aline Xavier Ferraz, Allan Abuabara, Flares Baratto-Filho, Bianca Marques de Mattos de Araujo, Erika Calvano Kuchler, Liliane Roskamp, Angela Graciela Deliga Schroder, and Cristiano Miranda de Araujo. 2025. Dental Trauma Evo–Development of an Artificial Intelligence-Powered Chatbot to Support Professional Management of Dental Trauma. *Journal of Endodontics* (2025).

[4] Tianlong Chen, Xuxi Chen, Xianzhi Du, Abdullah Rashwan, Fan Yang, Huizhong Chen, Zhangyang Wang, and Yeqing Li. 2023. Adamv-moe: Adaptive multi-task vision mixture-of-experts. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 17346–17357.

[5] Zhangquan Chen, Chunjiang Liu, and Haobin Duan. 2024. A Three-Phases-LORA Finetuned Hybrid LLM Integrated with Strong Prior Module in the Education Context. In *International Conference on Artificial Neural Networks*. Springer, 235–250.

[6] Susmita Das, Amara Tariq, Thiago Santos, Sai Sandeep Kantareddy, and Imon Banerjee. 2023. Recurrent neural networks (RNNs): architectures, training tricks, and introduction to influential research. *Machine learning for Brain disorders* (2023), 117–138.

[7] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems* 36 (2023), 10088–10115.

[8] Ma Dongbo, Sami Miniaoui, Li Fen, Sara A Althubiti, and Theyab R Alsenani. 2023. Intelligent chatbot interaction system capable for sentimental analysis using hybrid machine learning algorithms. *Information Processing & Management* 60, 5 (2023), 103440.

[9] Filippo Florindi, Pasquale Fedele, and Giovanna Maria Dimitri. 2024. A novel solution for the development of a sentimental analysis chatbot integrating ChatGPT. *Personal and Ubiquitous Computing* 28, 6 (2024), 947–960.

[10] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948* (2025).

[11] Jiaao He, Jidong Zhai, Tiago Antunes, Haojie Wang, Fuwen Luo, Shangfeng Shi, and Qin Li. 2022. Fastermoe: modeling and optimizing training of large-scale dynamic pre-trained models. In *Proceedings of the 27th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*. 120–134.

[12] Jiwoo Hong, Noah Lee, and James Thorne. 2024. Orpo: Monolithic preference optimization without reference model. *arXiv preprint arXiv:2403.07691* (2024).

[13] Jingcheng Hu, Yinmin Zhang, Qi Han, Daxin Jiang, Xiangyu Zhang, and Heung-Yeung Shum. 2025. Open-reasoner-zero: An open source approach to scaling up reinforcement learning on the base model. *arXiv preprint arXiv:2503.24290* (2025).

[14] Baha Ihnaini, Yawen Huang, Lianglin Li, Jiayi Wei, and Shengyi Qi. 2024. Enhancing Chinese Medical Diagnostic Chatbot through Supervised Fine-Tuning of Large Language Models. In *2024 6th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI)*. IEEE, 205–212.

[15] Raisa Islam and Owana Marzia Moushi. 2024. Gpt-4o: The cutting-edge advancement in multimodal llm. *Authorea Preprints* (2024).

[16] Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720* (2024).

[17] Katikapalli Subramanyam Kalyan. 2024. A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal* 6 (2024), 100048.

[18] Zeynep Karkiner, Begum Yaman, Begum Zengin, Feride Nursena Cavli, and Mustafa Sert. 2024. ParsyBot: chatbot for baskent university related FAQs. In *2024 IEEE 18th International Conference on Semantic Computing (ICSC)*. IEEE, 168–175.

[19] Feriel Khennouche, Youssef Elmir, Nabil Djebari, Larbi Boubchir, Abdelkader Laouid, and Ahcene Bounceur. 2024. Comparative Analysis and Application of Large Language Models on FAQ Chatbots. In *2024 International Conference on Computational Intelligence and Network Systems (CINS)*. IEEE, 1–6.

[20] Pranjal Kumar. 2024. Large language models (LLMs): survey, technical frameworks, and future challenges. *Artificial Intelligence Review* 57, 10 (2024), 260.

[21] FatimaEzzahra Laghrissi, Samira Douzi, Khadija Douzi, and Badr Hssina. 2021. Intrusion detection systems using long short-term memory (LSTM). *Journal of Big Data* 8, 1 (2021), 65.

[22] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437* (2024).

[23] Chengqi Lyu, Songyang Gao, Yuzhe Gu, Wenwei Zhang, Jianfei Gao, Kuikun Liu, Ziyi Wang, Shuaibin Li, Qian Zhao, Haian Huang, et al. 2025. Exploring the limit of outcome reward for learning mathematical reasoning. *arXiv preprint arXiv:2502.06781* (2025).

[24] Junyan Qiu and Yiping Yang. 2024. Training Large Language Models to Follow System Prompt with Self-Supervised Fine-tuning. In *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[25] K Rajasekaran, John Amose, G Preethika, S Sangamithrra, and G Gayathiri. 2024. Innovations in Dental Care: Chatbot-Driven Efficiency. In *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Vol. 1. IEEE, 852–857.

[26] Samyam Rajbhandari, Conglong Li, Zhewei Yao, Minjia Zhang, Reza Yazdani Aminabadi, Ammar Ahmad Awan, Jeff Rasley, and Yuxiong He. 2022. Deepspeed-moe: Advancing mixture-of-experts inference and training to power next-generation ai scale. In *International conference on machine learning*. PMLR, 18332–18346.

[27] Kethireddy Maheedhar Reddy and Radha Guha. 2023. Automatic text summarization for conversational chatbot. In *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*. IEEE, 1–7.

[28] Fillipe Barros Rodrigues, William Ferreira Giozza, Robson de Oliveira Albuquerque, and Luis Javier García Villalba. 2022. Natural language processing applied to forensics information extraction with transformers and graph visualization. *IEEE Transactions on Computational Social Systems* (2022).

[29] Dafne Itzel Rojas González, Josue Aaron Soriano Rivero, and Jatziri Hernandez Hernandez. 2024. Wrap Your Mind Around Education: Applying Hugging Face to a Chatbot with AI. In *International Conference on Disruptive Technologies, Tech Ethics and Artificial Intelligence*. Springer, 444–454.

[30] Abdullahi B Saka, Lukumon O Oyedele, Lukman A Akanbi, Sikiru A Ganiyu, Daniel WM Chan, and Sururah A Bello. 2023. Conversational artificial intelligence in the AEC industry: A review of present status, challenges and opportunities. *Advanced Engineering Informatics* 55 (2023), 101869.

[31] Suryani and Mustakim. 2024. An Intelligent Chatbot for Faculty Administration Using Bidirectional LSTM and Seq2Seq Architecture. In *2024 International Conference on Smart Computing, IoT and Machine Learning (SIML)*. IEEE, 226–231.

[32] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[33] Uday Vallabhaneni, Yatish Wutla, Tribhangin Dichpally, Venkata Rami Reddy Ch, Meghamsh Reddy Gone, and P Lalitha Kumari. 2024. Mining Mate: A Chat Bot for Navigating Mining Regulations Using LLM Models. In *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Vol. 1. IEEE, 888–892.

[34] Luping Wang, Sheng Chen, Linnan Jiang, Shu Pan, Runze Cai, Sen Yang, and Fei Yang. 2025. Parameter-efficient fine-tuning in large language models: a survey of methodologies. *Artificial Intelligence Review* 58, 8 (2025), 227.

[35] Sheng-Kai Wang, Wan-Lin You, and Shang-Pin Ma. 2022. Semi-automatic Chatbot Generation for Web APIs. In *International Computer Symposium*. Springer, 267–278.

[36] Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, et al. 2024. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark. *Advances in Neural Information Processing Systems* 37 (2024), 95266–95290.

[37] Qing Xia, Haotian Zhao, and Ming Liu. 2024. Prompt Engineering Approach Study for Supervised Fine-Tuned (SFT) Large Language Models (LLMs) in Spacecraft Fault Diagnosis. In *2024 3rd Conference on Fully Actuated System Theory and Applications (FASTA)*. IEEE, 819–824.

[38] Sukwon Yun, Inyoung Choi, Jie Peng, Yangfan Wu, Jingxuan Bao, Qiyiwen Zhang, Jiayi Xin, Qi Long, and Tianlong Chen. 2024. Flex-moe: Modeling arbitrary modality combination via the flexible mixture-of-experts. *Advances in Neural Information Processing Systems* 37 (2024), 98782–98805.

[39] Chenggang Zhao, Chengqi Deng, Chong Ruan, Damai Dai, Huazuo Gao, Jiashi Li, Liyue Zhang, Panpan Huang, Shangyan Zhou, Shirong Ma, et al. 2025. Insights into deepseek-v3: Scaling challenges and reflections on hardware for ai architectures. In *Proceedings of the 52nd Annual International Symposium on Computer Architecture*. 1731–1745.

[40] Shiyong Zheng, Zahrah Yahya, Lei Wang, Ruihang Zhang, and Azadeh Noori Hoshyar. 2023. Multiheaded deep learning chatbot for increasing production and marketing. *Information Processing & Management* 60, 5 (2023), 103446.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009