DentalGPT: Incentivizing Multimodal Reasoning in Dentistry

Zhenyang Cai[2†], Jiaming Zhang[1†], Junjie Zhao[5,6†], Ziyi Zeng[2], Yanchao Li[2]
Jingyi Liang[2], Junying Chen[2], Yunjin Yang[2], Jiajun You[2,4], Shuzhi Deng[1],
Ruiqi Xie[1], Yuanting Chen[1], Xiangyi Feng[4], Jianquan Li[4]
Liangyi Chen[3*], Junwen Wang[5*], Shan Jiang[1*], Benyou Wang[2,7,8,9*]
[1] Shenzhen Stomatology Hospital (Pingshan) of Southern Medical University
[2] The Chinese University of Hong Kong, Shenzhen [3] Peking University
[4] Freedom AI [5] Faculty of Dentistry, The University of Hong Kong
[6] Beijing Institute of Collaborative Innovation [7] National Health Data Institute, Shenzhen
[8] Shenzhen Loop Area Institute [9] Shenzhen Institute of Big Data

Abstract

Reliable interpretation of multimodal dental data is essential for automated oral healthcare, yet existing multimodal large language models (MLLMs) exhibit limited understanding of dental images. Although complex reasoning improves performance, its gains in dentistry are substantially smaller than in other medical domains, indicating that complex reasoning is not yet sufficiently incentivized for dental diagnosis, likely due to insufficient domain knowledge and limited reinforcement learning on dental questions. We present **DentalGPT**, a dentistry-specialized MLLM trained via staged multimodal alignment and reinforcement learning. By constructing the largest annotated multimodal dental dataset to date with over 120k images, multimodal alignment provides the necessary domain knowledge foundation to support and incentivize complex reasoning, which is further strengthened through reinforcement learning. Experiments on expert-annotated benchmarks and dental subsets of medical VQA benchmarks show that DentalGPT achieves superior performance on disease classification and dental VQA tasks, outperforming many state-of-the-art MLLMs despite its compact 7B parameter scale.

1 Introduction

Dental healthcare is an essential area of public health, yet the workload of dental professionals continues to increase each year (Centers for Disease Control and Prevention, 2024; Jiang et al., 2024; Zheng et al., 2025; Lilford et al., 2025). To support both clinicians and patients, multimodal large language models (MLLMs) (Liu et al., 2023; Bai et al., 2025b; Team et al., 2023; Chen et al., 2024d) capable of interactive communication through dialogue have recently attracted significant attention, offering new possibilities for intelligent dental care. However, despite their promising performance in general medical applications (Awadalla et al., 2023; Li et al., 2023; Wu et al., 2023, 2024a; Chen et al., 2024b; Su et al., 2025; Xu et al., 2025), current MLLMs still face notable limitations when dealing with more specialized medical imaging problems, such as images in dentistry.

In recent studies (Su et al., 2025; Bai et al., 2025a; Team et al., 2025b; Pan et al., 2025), enabling complex reasoning has yielded substantial gains on challenging multimodal benchmarks. Motivated by this, we evaluate mainstream MLLMs on a dentistry benchmark (Hao et al., 2025a) with and without complex reasoning mode, and observe consistent improvements when reasoning is enabled; however, these gains are markedly smaller than those reported in broader medical settings (Su et al., 2025; Zhang et al., 2025; Pan et al., 2025), suggesting that complex reasoning in dentistry is not yet sufficiently incentivized. Based on this observation, we hypothesize two challenges: 1) insufficient visual understanding of dental images, likely due to scarce dental VQA data, which narrows the model's reasoning perspectives and limits its exploration space for dental questions; and 2) limited reinforcement learning on dentistry-related questions, which constrains the acquisition of domain-specific reasoning patterns.

To address these challenges, we inject dentistry-related knowledge into MLLMs and incentivize complex reasoning. Specifically, we curate a large-scale dental dataset by combining online dental images with textual descriptions or labels and professionally annotated images from dental hospitals, resulting in over 120k images with detailed descriptions and task-specific QA pairs. The descriptions emphasize diagnostically relevant visual cues to improve multimodal alignment, while the QA data supports downstream dental tasks. After that, we apply a reinforcement learning stage using Group Relative Policy Optimization (GRPO) (Shao et al.,

[†]Equal Contribution. [*]Corresponding authors.

Findings of the Association for Computational Linguistics: ACL 2026, pages 2811–2829

2024) to encourage more explanatory reasoning for dental questions. This process yields **DentalGPT**, a dentistry-specific MLLM equipped with complex reasoning capabilities.

A comprehensive evaluation was conducted to assess DentalGPT's dental image analysis capability. We first evaluated question answering using existing dentistry-focused VQA benchmarks, and then assessed disease identification using professionally annotated benchmarks. Despite having only 7B parameters, DentalGPT outperforms existing MLLMs in dental image understanding and question answering, demonstrating strong efficiency and domain specialization. We further analyze the impact of the two training stages through detailed ablations. The multimodal alignment stage substantially enriches dental knowledge and improves performance across image analysis tasks. These gains are further amplified by reinforcement learning, leading to higher disease classification accuracy and more professional identification of diagnostically relevant visual cues.

In summary, our contributions are: (1) We introduce **DentalGPT**, a dentistry-specialized MLLM capable of fine-grained dental image understanding and multimodal complex reasoning. (2) We curate the largest multimodal dental dataset to date, comprising over 120k dental images with detailed annotations of diagnostically relevant visual features. (3) We show that multimodal alignment, combined with targeted reinforcement learning, is crucial for incentivizing complex reasoning in domain-specific MLLMs.

2 On Incentivizing Multimodal Reasoning in Dentistry

2.1 Observations

Dentistry relies on identifying lesion-related visual cues in dental images. Although multimodal complex reasoning has shown substantial gains across medical tasks (Zhang et al., 2025; Pan et al., 2025; Su et al., 2025), its effectiveness in dentistry remains unclear. We therefore evaluate leading MLLMs [1] with and without complex reasoning on MMOral-OPG-Bench (Hao et al., 2025a), the only publicly available multimodal dentistry benchmark. As shown in Figure 1, current MLLMs exhibit limited performance on dentistry-related multimodal tasks, indicating insufficient ability to perceive lesion-related cues in dental images. This is further supported by Appendix A, leading to the following observation.

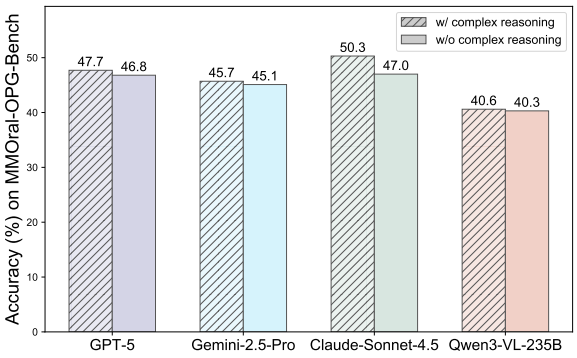


Figure 1: Accuracy (%) of MLLMs *with and without the complex reasoning mode* on the MMOral-OPG-Bench.

***Observation 1**: Current MLLMs struggle to reliably perceive lesion-related visual cues in dental images.*

Moreover, by comparing multiple MLLMs with and without the complex reasoning mode enabled, we further observe that:

***Observation 2**: Complex reasoning consistently improves performance, yet yields limited gains in dentistry.*

2.2 Analysis

Based on the above observations, we analyze why complex reasoning remains insufficiently incentivized in dentistry-related multimodal tasks. We group the potential reasons into two complementary levels: *perception* and *cognition*.

Perception-Level Limitation *Observation 1* suggests that insufficient multimodal alignment in dentistry limits visual grounding, narrows reasoning perspectives, and constrains the exploration space needed for incentivizing complex reasoning.

Cognition-Level Limitation *Observation 2* indicates that limited gains from complex reasoning stem from insufficient dental-specific optimization, as task-specific reinforcement learning for dentistry remains underexplored (Su et al., 2025; Zhang et al., 2025; Pan et al., 2025).

3 DentalGPT: From Multimodal Alignment to Complex Reasoning

To address perception- and cognition-level limitations in dental imaging, we propose a dentistry-

[0] **Complex reasoning mode:** GPT5-2025-08-07, Gemini-2.5-Pro-Thinking, Claude-Sonnet-4-5-20250929-Thinking, Qwen3-VL-235B-A22B-Thinking. **Without complex reasoning mode:** GPT5-chat-2025-08-07, Gemini-2.5-Pro-NoThinking, Claude-Sonnet-4-5-20250929, Qwen3-VL-235B-A22B-Instruct.

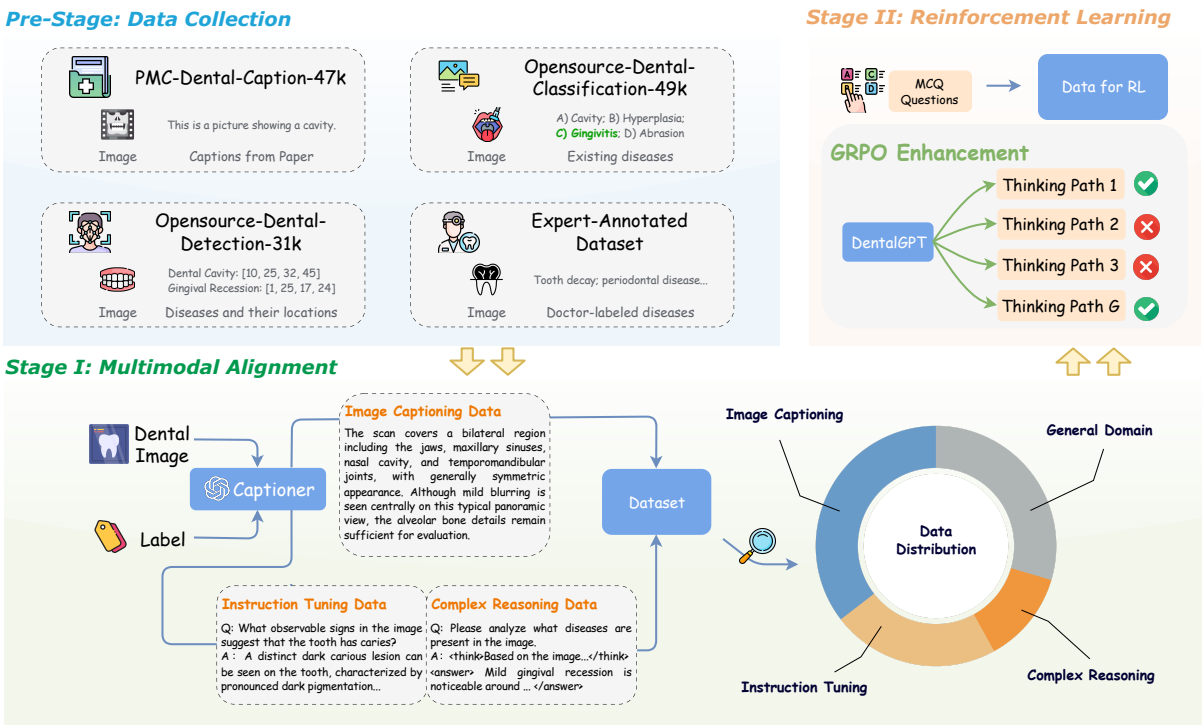


Figure 2: The process of building DentalGPT. Dentistry-related datasets from multiple sources are collected at the *Pre-stage: Data Collection*; *Stage I: Multimodal Alignment* uses the dataset to align the model's medical knowledge with its multimodal understanding; *Stage II: Reinforcement Learning* then strengthens complex reasoning ability.

specific framework to incentivize multimodal complex reasoning (Figure 2), with Stage I focusing on perception via multimodal alignment and Stage II on cognition via dentistry-oriented reinforcement learning.

3.1 Pre-stage: Data Collection

To support a comprehensive understanding of dental knowledge, diverse dental images were collected from multiple sources, including intraoral photographs, panoramic radiographs, and partial panoramics. These datasets provide complementary information that enhances the model's ability in dental knowledge learning, disease recognition, and spatial perception of lesions. Details of collection and filtering are provided in Appendix B.

Table 1 summarizes the dental image datasets used for multimodal training. PMC-Dental-Caption-47k provides rich image–text supervision from biomedical literature, enabling visual grounding and integration of dental domain knowledge. Opensource-Dental-Classification-49k supplies expert-labeled disease categories, supporting robust dental disease recognition. Opensource-Dental-Detection-31k further introduces spatial supervision through bounding box annotations, allowing the model to localize lesions, reason about spatial relationships, and quantify abnormalities. Together, these datasets offer complementary supervision signals for learning visual understanding, clinical semantics, and spatial reasoning in dentistry.

Expert-Annotated Dataset A subset of dental images was selected from internet sources, hospital imaging databases, and open-source datasets. After removing duplicates with the test set, professional dentists annotated these images with disease categories. Labels with high cross-validation agreement and sufficient sample sizes were retained for model evaluation.

3.2 Stage I: Multimodal Alignment

To address the perception-level limitations of existing MLLMs in dentistry, this stage enhances multimodal alignment by incorporating high-quality, domain-specific dental knowledge. We construct a large-scale, professionally curated image–text dataset to improve visual grounding and broaden perceptual coverage. `GPT-5` is used to generate detailed descriptions, strengthening fine-grained visual understanding and supporting basic downstream dental tasks.

Data Construction A large and comprehensive dataset was constructed, consisting of several components (Details can be found at Appendix C):

Dataset	Annotation Type	Learning Signal
PMC-Dental-Caption-47k	Image–text pairs from PubMed Central	Visual understanding with dental domain knowledge grounding
Opensource-Dental-Classification-49k	Expert-defined disease labels	Disease classification and key visual feature recognition
Opensource-Dental-Detection-31k	Bounding boxes for teeth and lesions	Spatial reasoning, localization, and lesion counting

Table 1: Overview of dental image datasets used for multimodal training, covering captioning, classification, and detection supervision.

- **Image Captioning** This component enhances the model's ability to capture diagnostically relevant visual details in dental images. `GPT-5` is instructed to describe all observable features that may aid diagnosis, while referencing original descriptions and labels and avoiding diagnostic assumptions. These observation-based captions are paired with predefined questions and answers to form a caption-based VQA subset, improving dental image interpretation and reducing visual information gaps.
- **Instruction Tuning** To improve performance on downstream tasks, `GPT-5` is prompted to generate questions from the collected images and descriptions, simulating real diagnostic scenarios. In addition, multiple vision models annotate a subset of the data, from which only high-confidence labels are retained. `GPT-5` then refines these annotations and converts them into structured question–answer pairs, further strengthening the model's dental question-answering capability.
- **Complex Reasoning** To support the subsequent post-training stage, complex chain-of-thought data were constructed by `GPT-5` following the `HuatuoGPT-o1` (Chen et al., 2024a) methodology. Fixed reasoning templates were added to activate the model's basic multi-step reasoning ability, enabling it to analyze before answering.
- **General Domain** General-domain and general medical data (Liu et al., 2023; An et al., 2025; Chen et al., 2024c) were included to preserve the model's broad visual understanding and prevent degradation of language capabilities during domain specialization.

After these construction steps were completed, `GPT-5-mini` was used to perform a secondary verification of all data. Entries that diverged from the original image descriptions or labels were removed, further ensuring data accuracy.

Training Settings The model was then trained on this dataset for two epochs with a batch size of 256 and a learning rate of 2×10^{-5}. All parameters were fully updated during training, with the first 5% of steps allocated for learning-rate warmup.

3.3 Stage II: Reinforcement Learning

After gaining new knowledge, the MLLM must learn to apply it for improved complex reasoning in multimodal diagnosis. Recent works such as `DeepSeek-R1` (DeepSeek-AI, 2025) and `GPT-o1` (Jaech et al., 2024) show that reinforcement learning can encourage long chain-of-thought generation and enhance reasoning quality. Following this paradigm, we adopt GRPO to optimize the reasoning process of *DentalGPT*.

Data Composition To achieve this goal, we selected a set of dental images that were not used during the Stage I to construct a new dataset. Based on the original labels and their label sets, we generated multiple-choice questions with correct answers, enabling rule-based correctness checking, which is crucial for reward computation in GRPO.

Training Strategy During this stage, we adopt the GRPO algorithm to improve the model's reasoning ability on dental multiple-choice tasks. GRPO optimizes relative advantages among sampled responses, enabling efficient training without a value network. We use answer accuracy as the primary reward and combine it with a format reward to enforce a fixed response template, weighted 90% and 10%, respectively.

Training Settings During GRPO optimization, the model was optimized using grouped rollouts with a sampling size of 10 responses per prompt. Training was conducted with a rollout batch size of 256 and a learning rate of 1×10^{-6}. The optimization ran for 5 epochs, and the maximum response length was capped at 8192 tokens to accommodate long CoT reasoning. This configuration ensured stable exploration within each action group while

maintaining sufficient capacity for detailed reasoning outputs.

4 Expert-annotated Benchmark

To comprehensively evaluate the model, a large set of dental images was collected and annotated with disease labels by professional dentists, ensuring clinical validity and allowing further assessment aligned with expert consensus.

4.1 Expert Annotation Workflow

Label Definition To maximize data diversity and ensure clinically reliable model outputs, we collected dental images from multiple sources and invited professional dentists to perform expert annotations. Guided by clinical practice, we defined a set of commonly observed dental diseases and diagnostically relevant signs that either directly indicate pathology or serve as auxiliary evidence in clinical reasoning. These labels are used to assess whether models can correctly identify key visual cues in dental images.

Cross-Validation To ensure annotation reliability, we implemented a rigorous cross-validation workflow. Each image was independently annotated by at least 2 dentists, who selected all clinically relevant labels present in the image. Annotators could mark images as *uncertain* when visual quality was insufficient, reducing forced or ambiguous judgments. Labels with any disagreement in cross-validation were removed, and only fully consistent labels were retained.

Benchmark Curation After expert annotation and quality filtering, only high-confidence and clinically reliable labels were retained to form the final dataset. This curated dataset serves as the foundation for constructing expert-annotated benchmarks and enables rigorous evaluation of multimodal diagnostic performance.

4.2 Benchmark Composition

Based on the above annotation workflow, we construct three benchmarks to evaluate the multimodal diagnostic performance of DentalGPT. These benchmarks cover both clinical and in-the-wild dental images across multiple modalities, enabling systematic assessment of reliability, generalization, and robustness in diverse dental scenarios. Each benchmark supports multi-label classification across both clinical and in-the-wild scenarios, enabling comprehensive evaluation of multimodal diagnostic ability. To avoid biased evaluation, we further applied strict data balancing strategies: the label distributions and the ratios of positive and negative samples for each category were aligned across subsets. This ensures that accuracy reliably reflect model performance rather than being influenced by label frequency or overrepresented diseases.

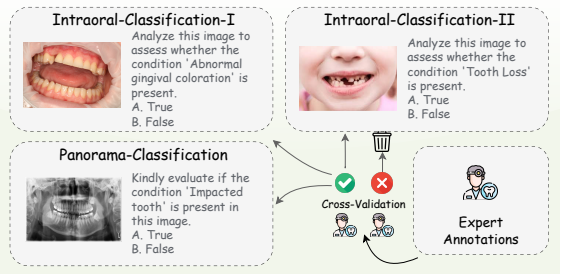


Figure 3: Examples of expert-annotated benchmarks. Only samples that pass cross-validation are retained.

Table 2 summarizes three dental image classification benchmarks with complementary evaluation settings. **Intraoral-Classification-I** uses dentist-acquired intraoral photographs collected under standardized clinical conditions and evaluates performance on high-quality, in-distribution images. **Intraoral-Classification-II** focuses on robustness and generalization, using patient-captured, in-the-wild images with diverse lighting and viewpoints. **Panorama-Classification** evaluates panoramic radiographs that reveal internal anatomical and pathological information beyond surface appearance. Together, these benchmarks assess clinical accuracy, robustness to real-world data, and cross-modality diagnostic capability.

5 Experiment

5.1 Evaluation of Dentistry Expertise

Training Settings DentalGPT is built upon `Qwen2.5-VL-7B-Instruct`. In both training stages, all model parameters are fully updated to enable domain adaptation for dentistry and complex reasoning. Detailed hyperparameters and implementation details are provided in Section 3. All experiments are conducted on a cluster with 8 NVIDIA H200 GPUs.

Evaluation Settings We conduct evaluations on the curated dentistry-specific benchmark (Section 4), which assesses models across dental disease classification, lesion recognition, and common dental consultation scenarios. All models are required to provide responses to the given tasks. For

Benchmark	Data Source and Key Characteristics	Label Set
Intraoral-Classification-I (300 samples)	Dentist-acquired intraoral photographs from AlphaDent (Sosnin et al., 2025), captured under standardized clinical conditions with consistent lighting and viewpoints, representing in-distribution clinical images.	Tooth discoloration; abnormal gingival coloration; gingival recession; dental caries; tooth pigmentation; tooth defect or loss; tooth loss; dental calculus; abnormal tooth morphology; abnormal gingival morphology.
Intraoral-Classification-II (206 samples)	Patient-captured intraoral images collected from the internet, exhibiting diverse lighting, resolution, and shooting angles, reflecting non-clinical, in-the-wild conditions and domain shift.	Tooth pigmentation; abnormal gingival coloration; dental calculus; tooth loss; dental caries; abnormal gingival morphology; gingival recession.
Panorama-Classification (156 samples)	Hospital-collected panoramic radiographs (X-ray), providing internal anatomical and pathological cues beyond surface appearance and requiring radiographic interpretation.	Periodontal disease; root canal treatment; tooth defect or loss; jawbone lesion; periapical lesion; impacted tooth.

Table 2: Three dental image classification benchmarks highlighting differences in acquisition setting (clinical vs. in-the-wild), image modality (photograph vs. radiograph), and clinical complexity.

Model	MMOral OPG-Bench	DentalBench Mixed	Intraoral Classification-I	Intraoral Classification-II	Panorama Classification	Avg.
Open-source MLLMs						
Deepseek-VL2 (Wu et al., 2024b)	39.1	22.6	51.1	59.4	55.1	45.5
Mistral-Large-2512 (AI, 2024)	41.9	48.2	50.7	58.0	44.2	48.6
Phi-4-Multimodal-Instruct (Abouelenin et al., 2025)	38.5	44.4	52.2	63.3	61.5	52.0
Ernie-4.5-VL-424B-A47B* (Baidu-ERNIE-Team, 2025)	45.0	51.4	58.1	65.1	44.9	52.9
Qwen3-VL-235B-A22B-Instruct (Bai et al., 2025a)	40.3	51.6	50.7	58.0	55.8	51.3
Gemma-3-27B-it (Team, 2025)	42.2	43.0	51.5	61.4	59.6	51.5
GLM-4.5v* (Team et al., 2025b)	45.7	51.4	54.8	64.7	54.5	54.2
Qwen3-VL-235B-A22B-Thinking* (Bai et al., 2025a)	40.6	51.6	56.7	65.7	60.3	55.0
LLaMA-4-Maverick (AI, 2025)	<u>51.4</u>	53.9	<u>61.1</u>	67.1	59.0	58.5
Proprietary MLLMs						
Claude-Sonnet-4.5 (Anthropic, 2025)	47.0	50.4	51.9	59.4	50.0	51.7
Claude-Sonnet-4.5-Thinking* (Anthropic, 2025)	50.3	53.9	55.2	66.7	55.8	56.4
Grok-4.1-Fast	47.1	52.2	57.0	65.2	62.2	56.7
Gemini-2.5-Pro-Thinking* (Comanici et al., 2025)	45.7	**57.4**	57.0	65.2	<u>64.1</u>	57.9
GPT-4.1 (OpenAI, 2024)	47.2	51.7	60.4	70.5	61.5	58.3
GPT-5* (OpenAI, 2025)	47.7	54.3	59.3	<u>71.0</u>	63.5	<u>59.2</u>
Medical MLLMs						
HuatuoGPT-Vision (Chen et al., 2024b)	36.3	38.5	55.2	51.2	53.8	47.0
MedGemma-4B-it (Sellergren et al., 2025)	47.6	39.8	49.6	50.4	60.9	49.7
MedGemma-27B-it (Sellergren et al., 2025)	50.6	41.8	52.6	56.2	61.0	52.4
DentalGPT and Its Backbone						
Qwen2.5-VL-7B-Instruct (Bai et al., 2025b)	27.0	46.1	48.8	61.8	50.0	46.7
DentalGPT*	**60.0**	<u>54.4</u>	**64.1**	**72.9**	**84.0**	**67.1**

Table 3: Accuracy (%) of MLLMs on Dental-related VQA Benchmarks. **Bold** indicates the best score; <u>underlines</u> marks the second-best. * indicates that the model has activated complex reasoning.

models marked with an asterisk (*), complex reasoning mode is enabled, while other models are instructed to directly output the correct option without additional reasoning steps.

Benchmarks We evaluate DentalGPT on both existing and newly constructed benchmarks to assess its multimodal diagnostic capability in dentistry. Existing benchmarks mainly measure general dental image understanding, including ***MMOral-OPG-Bench*** (Hao et al., 2025a), which contains panoramic dental X-ray images with expert annotations and for which the closed-ended test split is used, and ***DentalBench-Mixed***, constructed by filtering tooth-related samples from medical VQA benchmarks such as PMC-VQA (Zhang et al., 2023), OmniMedVQA (Hu et al., 2024), and MedXpertQA-MM (Zuo et al., 2025). We further introduce dentistry-specific benchmarks focused on common dental diseases and signs, enabling a more comprehensive evaluation in diagnostically critical scenarios.

Results and Analysis As shown in Table 3, DentalGPT consistently outperforms both comparable and substantially larger MLLMs across all expert-annotated benchmarks, achieving significant gains over its backbone and highlighting the effectiveness of the proposed data pipeline and domain-aligned training. Despite its compact 7B scale, DentalGPT matches or surpasses many general-purpose models exceeding 100B parameters. It also maintains

strong performance on *MMOral-OPG-Bench* and *DentalBench-Mixed*, demonstrating robust generalization across diverse dental tasks. Overall, these results position DentalGPT as a strong multimodal foundation model for dental image understanding.

5.2 Human Evaluation

To further assess DentalGPT's capabilities, four licensed dentists were invited to evaluate the model outputs based on predefined criteria (Details in Appendix D), focusing on obvious errors, the soundness of diagnoses, and omissions of critical information.

Specifically, 80 dental images were curated from medical textbooks and licensed dentist examinations, and each model was tasked with analyzing the images and identifying all potential dental conditions. DentalGPT was compared against several widely used general-purpose models through pairwise comparisons, and win/tie/loss statistics were reported. As shown in Figure 4, DentalGPT is preferred in most cases, with its outputs judged as more clinically accurate and better aligned with dentists. Notably, despite its smaller parameter scale, DentalGPT achieves performance comparable to leading models, offering a favorable trade-off between model size and clinical quality.

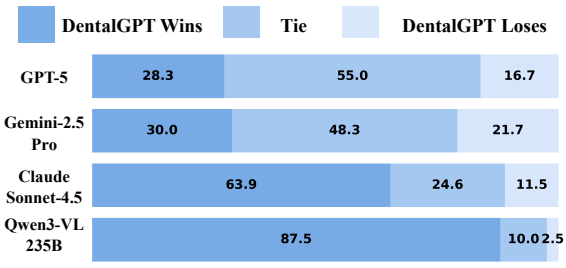


Figure 4: Results of the human evaluation. *Win/Tie/Loss* indicates the ratios of expert preferences on responses.

5.3 Ablation Study of Multimodal Alignment (on Stage I)

Multimodal alignment equips MLLMs with a richer understanding of dentistry-related images. In this subsection, we examine whether such alignment facilitates subsequent complex reasoning.

To characterize changes in complex reasoning, we analyze the evolution of the accuracy-based reward during reinforcement learning after alignment. In this process, the model repeatedly relies on its internalized knowledge to explore different solution paths for improving downstream task accuracy. Consequently, the accuracy reward serves as an indicator of whether the injected alignment knowledge supports more diverse exploration and more elaborate reasoning behaviors, thereby facilitating complex multimodal reasoning.

Experimental Setup Specifically, three controlled settings using 0%, 30%, and 100% of the Stage-I dataset were evaluated. The impact of alignment strength was assessed by tracking accuracy-based reward improvements during the subsequent RL stage. To ensure fairness, duplicate images between alignment and RL data were removed, and all *complex reasoning* samples were excluded from Stage I. Following the DentalGPT setup, `Qwen2.5-VL-7B` was adopted as the backbone, and each model was trained for 30 RL steps, with reward trajectories on the validation set used to assess multimodal reasoning performance.

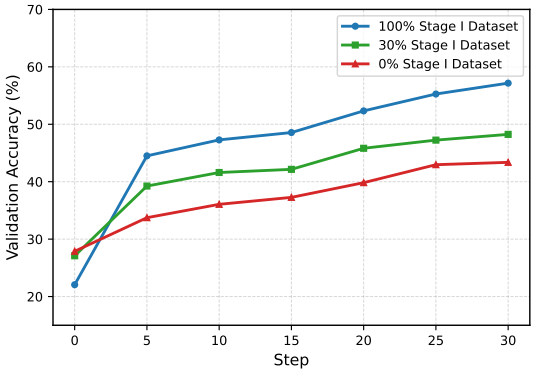


Figure 5: Accuracy reward (%) of the MLLM during RL training under different Stage I dataset scales, showing that stronger multimodal alignment facilitates more effective reward optimization.

Results Analysis As shown in Figure 5, the model trained with 0% Stage-I data exhibits only marginal reward gains during RL, indicating limited reasoning improvement without dental-domain alignment. In contrast, increasing the amount of Stage-I data consistently elevates the reward ceiling throughout RL training. These results suggest that *multimodal alignment provides the model with richer perspectives for exploration during the RL stage, thereby further incentivizing complex multimodal reasoning in dentistry.*

5.4 Ablation Study of RL (on Stage II)

This section investigates how the reinforcement learning (RL) stage further shapes the model's capabilities. After applying RL training on 10k multiple-choice dental questions, the performance of the model was reevaluated across the same set of

benchmarks. As shown in Table 4, reinforcement learning brings consistent improvements across all tasks, demonstrating that it further enhances the model's ability to execute downstream dental tasks. These gains confirm that *RL effectively strengthens both the accuracy and reliability of the model's reasoning in dental image understanding.*

Benchmarks	Qwen2.5-VL Backbone	Qwen2.5-VL + Stage I w/o Stage II	DentalGPT w/ Stage I & Stage II
MMOral-OPG-Bench	27.0	56.8	**60.0**
DentalBench-Mixed	46.1	51.7	**54.4**
Intraoral-Classification-I	48.8	61.5	**64.1**
Intraoral-Classification-II	61.8	67.6	**72.9**
Panorama-Classification	50.0	78.4	**84.0**
Total	46.7	63.2	**67.1**

Table 4: Accuracy(%) comparison among `Qwen2.5-VL`, `Qwen2.5-VL` with Stage I training, and DentalGPT (Stages I + II) on dentistry-specific benchmarks.

5.5 Case Study

We further analyze the outputs of **DentalGPT** across different training stages together with the original backbone model (Figure 6). The backbone model struggles with this example, describing some relevant visual features but failing to identify any teeth with fillings. After Stage-I multimodal alignment, the model can perform basic dental image interpretation and detect most filled teeth, missing only a less salient case, highlighting the role of multimodal alignment in enabling visual grounding and establishing a foundation for reasoning. After completing all training stages, **DentalGPT** with complex reasoning conducts a more systematic multimodal analysis and, despite intermediate inaccuracies, iteratively verifies and refines its predictions to arrive at the correct answer. This case study underscores the importance of reinforcement learning in supporting self-correction and accurate complex multimodal reasoning in dentistry.

6 Related Work

Medical MLLMs Medical imaging is a key application area for MLLMs due to its reliance on visual data. General medical MLLMs (Chen et al., 2024b; Lin et al., 2025; Team et al., 2025a; Chao et al., 2024; Yu et al., 2025) demonstrate basic medical QA and preliminary visual analysis across diverse clinical scenarios, emphasizing broad generalization through large-scale multimodal medical data and benchmarks (Zhang et al., 2023; Liu et al., 2021; Xie et al., 2024; Cai et al., 2025).

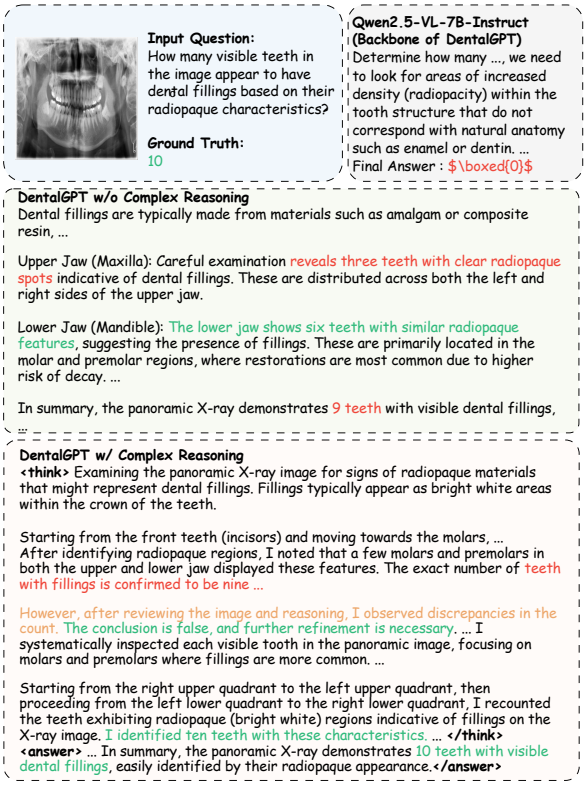


Figure 6: Case study of **DentalGPT** and its backbone on a dental multimodal task. Red and green indicate incorrect and correct analyses, while orange highlights reflective reasoning.

MLLMs for Specific Medical Domains To address complex clinical settings, recent work adapts MLLMs to specific medical modalities, including 3D medical imaging (Wu et al., 2023, 2025b,a), biomedical signal understanding (Zhao et al., 2024; Liu et al., 2024; Chen et al., 2025b; Zeng et al., 2025), and fine-grained perception of high-resolution clinical images via specialized pipelines or training strategies (Chen et al., 2025a; Wang et al., 2025).

MLLMs for Dentistry Dentistry is an important and emerging multimodal domain for medical AI. In this context, `DentVLM` (Meng et al., 2025) leverages large-scale dental reports to support dental tasks, while `OralGPT` (Hao et al., 2025a,b) introduces dentistry-focused multimodal benchmarks and enables diverse dental modalities. We hope that our work and theirs can jointly contribute to the advancement of intelligent AI in dentistry.

7 Conclusion

This work presents **DentalGPT**, a specialized MLLM for multimodal dental diagnosis. By constructing a large-scale multimodal dataset in den-

tistry and adopting a staged training strategy that combines multimodal alignment and reinforcement learning, DentalGPT effectively captures fine-grained visual cues and supports reliable disease-related reasoning. Extensive evaluations across different dental VQA benchmarks show that DentalGPT achieves strong performance with only 7B parameters, outperforming many state-of-the-art general-purpose MLLMs.

Limitations

Despite extensive data collection efforts, the availability of dental imaging data remains limited, especially for high-quality annotated datasets. As a result, our data construction relies on expert annotation and human-in-the-loop curation, which is costly and difficult to scale. Future work may explore more efficient human–AI collaboration and improved data validation to expand dataset scale while preserving quality.

Potential Risks

Although DentalGPT shows strong performance and high preference among licensed dentists, it is not intended for direct clinical use. Real-world deployment would require further adaptation, rigorous validation, and integration into clinical workflows to ensure safety and reliability. Clinical decisions should remain under the supervision of qualified professionals.

Acknowledgement

This work was supported by Shenzhen Medical Research Fund (B2503005), Major Frontier Exploration Program (Grant No. C10120250085) from the Shenzhen Medical Academy of Research and Translation (SMART), the Shenzhen Science and Technology Program (JCYJ20220818103001002), NSFC grant 72495131, Shenzhen Doctoral Startup Funding (RCBS20221008093330065), Tianyuan Fund for Mathematics of National Natural Science Foundation of China (NSFC) (12326608), Shenzhen Science and Technology Program (Shenzhen Key Laboratory Grant No. ZDSYS20230626091302006), the 1+1+1 CUHK-CUHK(SZ)-GDSTC Joint Collaboration Fund, Guangdong Provincial Key Laboratory of Mathematical Foundations for Artificial Intelligence (2023B1212010001), Shenzhen Stability Science Program 2023, and the Collaborative Research Fund of the Hong Kong Research Grants Council (RGC) (C7015-23G, to JZ and JW).

References

Abdelrahman Abouelenin, Atabak Ashfaq, Adam Atkinson, Hany Awadalla, Nguyen Bach, Jianmin Bao, Alon Benhaim, Martin Cai, Vishrav Chaudhary, Congcong Chen, and 1 others. 2025. Phi-4-mini technical report: Compact yet powerful multimodal language models via mixture-of-loras. *arXiv preprint arXiv:2503.01743*.

Meta AI. 2025. Llama 4 maverick. https://www.llama.com/models/llama-4/. Accessed: 2025-12-06.

Mistral AI. 2024. Mistral large instruct 2407. https://huggingface.co/mistralai/Mistral-Large-Instruct-2407. Accessed: 2025-12-06.

Xiang An, Yin Xie, Kaicheng Yang, Wenkang Zhang, Xiuwei Zhao, Zheng Cheng, Yirui Wang, Songcen Xu, Changrui Chen, Chunsheng Wu, Huajie Tan, Chunyuan Li, Jing Yang, Jie Yu, Xiyao Wang, Bin Qin, Yumeng Wang, Zizhen Yan, Ziyong Feng, and 3 others. 2025. Llava-onevision-1.5: Fully open framework for democratized multimodal training. In *arXiv*.

Anthropic. 2025. Claude opus 4-5 system card. https://assets.anthropic.com/m/64823ba7485345a7/Claude-Opus-4-5-System-Card.pdf. Accessed: 2025-11-30.

Anas Awadalla, Irena Gao, Josh Gardner, Jack Hessel, Yusuf Hanafy, Wanrong Zhu, Kalyani Marathe, Yonatan Bitton, Samir Gadre, Shiori Sagawa, and 1 others. 2023. Openflamingo: An open-source framework for training large autoregressive vision-language models. *arXiv preprint arXiv:2308.01390*.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, and 1 others. 2025a. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631*.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, and 1 others. 2025b. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*.

Baidu-ERNIE-Team. 2025. Ernie 4.5 technical report. https://ernie.baidu.com/blog/publication/ERNIE_Technical_Report.pdf.

Zhenyang Cai, Junying Chen, Rongsheng Wang, Weihong Wang, Yonglin Deng, Dingjie Song, Yize Chen, Zixu Zhang, and Benyou Wang. 2025. Exploring compositional generalization of multimodal llms for medical imaging. In *Proceedings of the 63rd Annual Meeting of the Association for Computational*

Linguistics (Volume 1: Long Papers), pages 13057–13079.

Centers for Disease Control and Prevention. 2024. 2024 oral health surveillance report: Selected findings. Accessed: 2025-04-28.

Chieh-Ju Chao, Imon Banerjee, Reza Arsanjani, Chadi Ayoub, Andrew Tseng, Garvan C. Kane, Jae K Oh, Li Fei-Fei, Ehsan Adeli, and Curtis Langlotz. 2024. Echogpt: A large language model for echocardiography report summarization. *medRxiv*.

Jingyun Chen, Linghan Cai, Zhikang Wang, Yi Huang, Songhan Jiang, Shenjin Huang, Hongpeng Wang, and Yongbing Zhang. 2025a. Pathagent: Toward interpretable analysis of whole-slide pathology images via large language model-based agentic reasoning. *arXiv preprint arXiv:2511.17052*.

Junying Chen, Zhenyang Cai, Ke Ji, Xidong Wang, Wanlong Liu, Rongsheng Wang, Jianye Hou, and Benyou Wang. 2024a. Huatuogpt-o1, towards medical complex reasoning with llms. *arXiv preprint arXiv:2412.18925*.

Junying Chen, Zhenyang Cai, Zhiheng Liu, Yunjin Yang, Rongsheng Wang, Qingying Xiao, Xiangyi Feng, Zhan Su, Jing Guo, Xiang Wan, and 1 others. 2025b. Shizhengpt: Towards multimodal llms for traditional chinese medicine. *arXiv preprint arXiv:2508.14706*.

Junying Chen, Chi Gui, Ruyi Ouyang, Anningzhe Gao, Shunian Chen, Guiming Hardy Chen, Xidong Wang, Ruifei Zhang, Zhenyang Cai, Ke Ji, Guangjun Yu, Xiang Wan, and Benyou Wang. 2024b. Huatuogpt-vision, towards injecting medical visual knowledge into multimodal llms at scale. *Preprint*, arXiv:2406.19280.

Lin Chen, Jinsong Li, Xiaoyi Dong, Pan Zhang, Conghui He, Jiaqi Wang, Feng Zhao, and Dahua Lin. 2024c. Sharegpt4v: Improving large multi-modal models with better captions. In *European Conference on Computer Vision*, pages 370–387. Springer.

Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, and 1 others. 2024d. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24185–24198.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, and 1 others. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*.

DeepSeek-AI. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

Jing Hao, Yuxuan Fan, Yanpeng Sun, Kaixin Guo, Lizhuo Lin, Jinrong Yang, Qi Yong H Ai, Lun M Wong, Hao Tang, and Kuo Feng Hung. 2025a. Towards better dental ai: A multimodal benchmark and instruction dataset for panoramic x-ray analysis. *arXiv preprint arXiv:2509.09254*.

Jing Hao, Yuci Liang, Lizhuo Lin, Yuxuan Fan, Wenkai Zhou, Kaixin Guo, Zanting Ye, Yanpeng Sun, Xinyu Zhang, Yanqi Yang, and 1 others. 2025b. Oralgpt-omni: A versatile dental multimodal large language model. *arXiv preprint arXiv:2511.22055*.

Yutao Hu, Tianbin Li, Quanfeng Lu, Wenqi Shao, Junjun He, Yu Qiao, and Ping Luo. 2024. Omnimedvqa: A new large-scale comprehensive evaluation benchmark for medical lvlm. *arXiv preprint arXiv:2402.09181*.

Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, and 1 others. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720*.

Xiaochen Jiang, Zhiguo Ding, Yanlei Su, Fei Wang, Weifeng Wang, Ziyang Wang, Xueling Qiu, Chenxi Sun, Fan Sun, Lu Tang, and 1 others. 2024. Dentists' views on the role orientation of dental hygienists in china: A qualitative content analysis. *BMC Oral Health*, 24(1):1563.

Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. 2023. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *arXiv preprint arXiv:2306.00890*.

Richard J Lilford, Benjamin Daniels, Barbara McPake, Zulfiqar A Bhutta, Robert Mash, Frances Griffiths, Akinyinka Omigbodun, Elzo Pereira Pinto, Radhika Jain, Gershim Asiki, and 1 others. 2025. Supply-side and demand-side factors affecting allopathic primary care service delivery in low-income and middle-income country cities. *The Lancet Global Health*, 13(5):e942–e953.

Tianwei Lin, Wenqiao Zhang, Sijing Li, Yuqian Yuan, Binhe Yu, Haoyuan Li, Wanggui He, Hao Jiang, Mengze Li, Xiaohui Song, Siliang Tang, Jun Xiao, Hui Lin, Yueting Zhuang, and Beng Chin Ooi. 2025. Healthgpt: A medical large vision-language model for unifying comprehension and generation via heterogeneous knowledge adaptation. *Preprint*, arXiv:2502.09838.

Bo Liu, Li-Ming Zhan, Li Xu, Lin Ma, Yan Yang, and Xiao-Ming Wu. 2021. Slake: A semantically-labeled knowledge-enhanced dataset for medical visual question answering. In *2021 IEEE 18th international symposium on biomedical imaging (ISBI)*, pages 1650–1654. IEEE.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual instruction tuning. *Advances in*

neural information processing systems, 36:34892–34916.

Ruoqi Liu, Yuelin Bai, Xiang Yue, and Ping Zhang. 2024. Teach multimodal llms to comprehend electrocardiographic images. *arXiv preprint arXiv:2410.19008*.

Zijie Meng, Jin Hao, Xiwei Dai, Yang Feng, Jiaxiang Liu, Bin Feng, Huikai Wu, Xiaotang Gai, Hengchuan Zhu, Tianxiang Hu, and 1 others. 2025. Dentvlm: A multimodal vision-language model for comprehensive dental diagnosis and enhanced clinical practice. *arXiv preprint arXiv:2509.23344*.

OpenAI. 2024. Gpt-4.1. https://openai.com/index/gpt-4-1/. Accessed: 2024-11-30.

OpenAI. 2025. Gpt-5 system card. Available at https://cdn.openai.com/gpt-5-system-card.pdf.

Jiazhen Pan, Che Liu, Junde Wu, Fenglin Liu, Jiayuan Zhu, Hongwei Bran Li, Chen Chen, Cheng Ouyang, and Daniel Rueckert. 2025. Medvlm-r1: Incentivizing medical reasoning capability of vision-language models (vlms) via reinforcement learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 337–347. Springer.

Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, Atilla Kiraly, Madeleine Traverse, Timo Kohlberger, Shawn Xu, Fayaz Jamil, Cían Hughes, Charles Lau, and 1 others. 2025. Medgemma technical report. *arXiv preprint arXiv:2507.05201*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *Preprint*, arXiv:2402.03300.

Evgeniy I. Sosnin, Yuriy L. Vasilev, Roman A. Solovyev, Aleksandr L. Stempkovskiy, Dmitry V. Telpukhov, Artem A. Vasilev, Aleksandr A. Amerikanov, and Aleksandr Y. Romanov. 2025. Alphadent: A dataset for automated tooth pathology detection. *Preprint*, arXiv:2507.22512.

Yanzhou Su, Tianbin Li, Jiyao Liu, Chenglong Ma, Junzhi Ning, Cheng Tang, Sibo Ju, Jin Ye, Pengcheng Chen, Ming Hu, and 1 others. 2025. Gmai-vl-r1: Harnessing reinforcement learning for multimodal medical reasoning. *arXiv preprint arXiv:2504.01886*.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, and 1 others. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Gemma Team. 2025. Gemma 3. *arXiv preprint arXiv:2503.19786*.

LASA Team, Weiwen Xu, Hou Pong Chan, Long Li, Mahani Aljunied, Ruifeng Yuan, Jianyu Wang, Chenghao Xiao, Guizhen Chen, Chaoqun Liu, Zhaodonghui Li, Yu Sun, Junao Shen, Chaojun Wang, Jie Tan, Deli Zhao, Tingyang Xu, Hao Zhang, and Yu Rong. 2025a. Lingshu: A generalist foundation model for unified multimodal medical understanding and reasoning. *Preprint*, arXiv:2506.07044.

V Team, Wenyi Hong, Wenmeng Yu, Xiaotao Gu, Guo Wang, Guobing Gan, Haomiao Tang, Jiale Cheng, Ji Qi, Junhui Ji, Lihang Pan, Shuaiqi Duan, Weihan Wang, Yan Wang, Yean Cheng, Zehai He, Zhe Su, Zhen Yang, Ziyang Pan, and 69 others. 2025b. Glm-4.5v and glm-4.1v-thinking: Towards versatile multimodal reasoning with scalable reinforcement learning. *Preprint*, arXiv:2507.01006.

Sheng Wang, Ruiming Wu, Charles Herndon, Yihang Liu, Shunsuke Koga, Jeanne Shen, and Zhi Huang. 2025. Pathology-cot: Learning visual chain-of-thought agent from expert whole slide image diagnosis behavior. *arXiv preprint arXiv:2510.04587*.

Chaoyi Wu, Weixiong Lin, Xiaoman Zhang, Ya Zhang, Weidi Xie, and Yanfeng Wang. 2024a. Pmc-llama: toward building open-source language models for medicine. *Journal of the American Medical Informatics Association*, page ocae045.

Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Hui Hui, Yanfeng Wang, and Weidi Xie. 2025a. Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data. *Nature Communications*, 16(1):7866.

Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. 2023. Towards generalist foundation model for radiology. *arXiv preprint arXiv:2308.02463*.

Jing Wu, Yuli Wang, Zhusi Zhong, Weihua Liao, Natalia Trayanova, Zhicheng Jiao, and Harrison X Bai. 2025b. Vision-language foundation model for 3d medical imaging. *npj Artificial Intelligence*, 1(1):17.

Zhiyu Wu, Xiaokang Chen, Zizheng Pan, Xingchao Liu, Wen Liu, Damai Dai, Huazuo Gao, Yiyang Ma, Chengyue Wu, Bingxuan Wang, and 1 others. 2024b. Deepseek-vl2: Mixture-of-experts vision-language models for advanced multimodal understanding. *arXiv preprint arXiv:2412.10302*.

Yunfei Xie, Ce Zhou, Lang Gao, Juncheng Wu, Xianhang Li, Hong-Yu Zhou, Sheng Liu, Lei Xing, James Zou, Cihang Xie, and 1 others. 2024. Medtrinity-25m: A large-scale multimodal dataset with multigranular annotations for medicine. *arXiv preprint arXiv:2408.02900*.

Weiwen Xu, Hou Pong Chan, Long Li, Mahani Aljunied, Ruifeng Yuan, Jianyu Wang, Chenghao Xiao, Guizhen Chen, Chaoqun Liu, Zhaodonghui Li, and 1 others. 2025. Lingshu: A generalist foundation model for unified multimodal medical understanding and reasoning. *arXiv preprint arXiv:2506.07044*.

Yuechun Yu, Han Ying, Haoan Jin, Wenjian Jiang, Dong Xian, Binghao Wang, Zhou Yang, and Mengyue Wu. 2025. Medkgeval: A knowledge graph-based multi-turn evaluation framework for open-ended patient interactions with clinical llms. *Preprint*, arXiv:2510.12224.

Ziyi Zeng, Zhenyang Cai, Yixi Cai, Xidong Wang, Junying Chen, Rongsheng Wang, Yipeng Liu, Siqi Cai, Benyou Wang, Zhiguo Zhang, and 1 others. 2025. Wavemind: Towards a conversational eeg foundation model aligned to textual and visual modalities. *arXiv preprint arXiv:2510.00032*.

Sheng Zhang, Qianchu Liu, Guanghui Qin, Tristan Naumann, and Hoifung Poon. 2025. Medrlvr: Emerging medical reasoning from a 3b base model via reinforcement learning. *arXiv preprint arXiv:2502.19655*.

Xiaoman Zhang, Chaoyi Wu, Ziheng Zhao, Weixiong Lin, Ya Zhang, Yanfeng Wang, and Weidi Xie. 2023. Pmc-vqa: Visual instruction tuning for medical visual question answering. *arXiv preprint arXiv:2305.10415*.

Yubao Zhao, Tian Zhang, Xu Wang, Puyu Han, Tong Chen, Linlin Huang, Youzhu Jin, and Jiaju Kang. 2024. Ecg-chat: A large ecg-language model for cardiac disease diagnosis. *arXiv preprint arXiv:2408.08849*.

Peixin Zheng, Xiaoting Qiu, Lingxiao Zhang, Peizhang Liu, Zeyi Peng, and Zhijian Huang. 2025. Comparative analysis of oral disorder burden in china and globally from 1990 to 2021 based on gbd data. *Scientific Reports*, 15(1):10061.

Yuxin Zuo, Shang Qu, Yifei Li, Zhangren Chen, Xuekai Zhu, Ermo Hua, Kaiyan Zhang, Ning Ding, and Bowen Zhou. 2025. Medxpertqa: Benchmarking expert-level medical reasoning and understanding. *arXiv preprint arXiv:2501.18362*.

A Case Study

Dentistry is a key medical field that relies on clinicians analyzing patients' imaging data and communicating with them for diagnosis, yet even leading MLLMs with strong general multimodal abilities still fall far behind professional dentists in multimodal diagnosis.

To investigate this phenomenon, we conducted a case study on a dental multimodal task using one leading commercial MLLM as well as both the *complex reasoning* and *non-complex reasoning* modes of a leading MLLM (Figure 7) [2]. After analyzing the model outputs, we found that although they could identify the relevant visual features to be counted, they still produced incorrect answers. Notably, by examining the reasoning trajectory of `Qwen3-VL-235B-A22B-Thinking`, we observed that it repeatedly reflected on and revised its own intermediate conclusions. Through this type of complex reasoning pattern, the model gradually approached the correct answer step by step. Although it still failed to produce the correct final prediction, this behavior highlights the potential of applying complex reasoning strategies to achieve more accurate multimodal diagnosis in dentistry.

[2] `Qwen3-VL-235B-A22B` is a fully open-source model family with both Thinking and Instruct models, corresponding to the with and without complex reasoning modes. Its fully accessible reasoning paths and diverse model variants make it the primary model used in this study.

B Data Collection

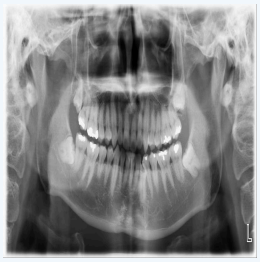
Before improving an MLLM's understanding and reasoning on dental images, it is necessary to collect a sufficiently large set of training samples, which provides the foundation for the model to effectively use its knowledge during image interpretation (Liu et al., 2023).

B.1 Existing Annotated Data

To efficiently obtain a sufficiently large number of dentistry-specific multimodal datasets, we first sourced datasets from a variety of open-source platforms. By leveraging certified, high-quality dental image datasets and existing literature, we aim to enrich the model's understanding of relevant radiographic data.

PMC-Dental-Caption-47k PubMed Central (PMC) is a publicly accessible biomedical repository that hosts a vast collection of peer-reviewed medical publications. It is considered a reliable and widely used data source in previous research. From PMC, we filtered a large number of dental images and retained the associated captions and labels provided within the original papers. This rich textual context is expected to provide valuable information for enhancing visual understanding and facilitating the integration of dental domain knowledge into our model.

Opensource-Dental-Classification-49k To further leverage image datasets that have previously been used to train classification models, we collected a wide range of dental-related classification datasets and consolidated them into a larger corpus of dental images with corresponding labels. Specifically, each image is associated with one or more disease labels; for multi-class or multi-label datasets, we also retain the negative labels so that



Input Question:
How many visible teeth in the image appear to have dental fillings based on their radiopaque characteristics?

Ground Truth:
10

GPT-5
From the panoramic X-ray, the radiopaque (bright white) areas on the chewing surfaces of several molars indicate dental fillings...
By visually examining the image, there appear to be: Approximate number of teeth with fillings: 6 teeth. ...

Qwen3-VL-235B-A22B-Instruct
Based on the provided panoramic dental radiograph, we can identify dental fillings by their radiopaque (bright white) appearance within the tooth structure — typically in the occlusal, proximal, or cervical areas — contrasting with the less dense enamel and dentin. Let's analyze the image systematically: Upper Arch ... Lower Arch ...

Total: 8 (upper)+8 (lower)=16 teeth with visible radiopaque fillings ...

Qwen3-VL-235B-A22B-Thinking
<think> Okay, let's see. ... First, I need to remember that in dental X-rays, fillings (amalgam or other materials) appear radiopaque, meaning they show up as white or bright areas. So I need to count the teeth with those ...

... There are 4 teeth with visible radiopaque fillings ... Wait, but maybe more. Let me check the image again. In the lower jaw, looking at the molars: ... There are 6 teeth with visible radiopaque fillings ...

But the question says "appear to have dental fillings based on radiopaque characteristics". So if it's visible, even if small. Let me count again: ... Total 8. Yes, I think that's it! **</think>**
<answer> ... Total count: 8 teeth ...**</answer>**

Figure 7: Examples of top-tier general-purpose MLLMs analyzing a dental image task. Red indicates incorrect analysis, green indicates correct analysis, and orange highlights reflective turns in the complex reasoning process.

all available, clinically validated annotations can be fully exploited. This unified resource supports MLLMs in better aligning common dental disease categories with key feature identification.

Opensource-Dental-Detection-31k We also collected a number of datasets previously utilized for dental lesion localization tasks, in which each image is annotated with one or more lesion instances together with their spatial coordinates. Although our model is not explicitly trained to predict bounding boxes, such annotations provide MLLMs with implicit spatial cues and lesion counts, thereby supporting the model's capability to understand spatial relationships and quantify dental abnormalities.

B.2 Newly Annotated Data

Building upon the open-source datasets described above, we observed that although they contain a considerable number of annotations, the diagnostic focus is predominantly centered on a few common dental conditions. From a clinical perspective, however, there exist additional critical abnormalities and visual manifestations that warrant greater attention yet are underrepresented in existing data sources. To address this gap, we expanded the diagnostic label set and further curated a new subset of dental images, which were annotated by certified dental experts with an emphasis on clinically significant conditions and indicative visual features.

Expert-Annotated Dataset We collected some dental images from internet sources, hospital imaging archives, and publicly available datasets. After removing duplicates and low-resolution images, we curated a candidate dataset that was subsequently annotated by professional dental clinicians. To ensure annotation quality, a strict cross-validation mechanism was applied with different levels of control over data variation. For the training set, annotations from dentists with a cross-validation agreement rate below 85% were filtered out, while the remaining labels were retained for constructing caption-style training data. For the test set, a more rigorous protocol was adopted: each sample was annotated by at least two dentists, and only those with consistent diagnostic results were preserved. This dataset equips the model with specialized clinical knowledge and a broader ability to recognize dental diseases and clinically relevant signs.

C Data Curation

C.1 Curation Strategies

To enrich the model's multimodal alignment, instruction-following ability, and complex reasoning in dentistry, we further curate and refine all collected data using `GPT-5`. This data curation process is designed to provide high-quality supervision for both fine-grained dental image understanding and the elicitation of multi-step reasoning, serving as a foundation for subsequent reinforcement learning. The overall process consists of image captioning, instruction tuning, and complex reasoning data construction, as outlined below.

Image Captioning This component focuses on enhancing the model's ability to capture diagnostically relevant visual details in dental images. Given a dental image $\mathcal{I}$, a reference label l, and a captioning instruction q_{cap}, the LLM is prompted to generate an observation-based description d that enumerates all visible cues related to the label while avoiding diagnostic assumptions (Detailed prompts are in Appendix E):

$$d = \text{LLM}(\mathcal{I}, q_{\text{cap}}, l).$$

These captions are then paired with predefined dental questions and answers to form caption-based VQA samples, allowing the model to better associate fine-grained visual evidence with corresponding textual concepts and reducing visual information gaps.

Instruction Tuning To improve performance on downstream dental tasks, we further construct instruction-following supervision using a two-step generation process. In the first step, given the dental image $\mathcal{I}$, an instruction q_{inst}, and the reference label l, the LLM generates an intermediate question–answer pair (q^*, a^*) that reflects a plausible dental diagnostic query:

$$(q^*, a^*) = \text{LLM}(\mathcal{I}, q_{\text{inst}}, l).$$

In the second step, the LLM produces a final answer a conditioned on the image, the instruction, and the generated intermediate pair:

$$a = \text{LLM}(\mathcal{I}, q_{\text{inst}}, q^*, a^*).$$

In addition, several vision models are used to annotate a subset of images, and only high-confidence predictions are retained. The LLM further refines these annotations and converts them into structured question–answer pairs, which are incorporated into the instruction tuning data to improve dental question answering under diverse task formulations.

Complex Reasoning To support the subsequent reinforcement learning stage, we construct complex chain-of-thought (CoT) data following the methodology of `HuatuoGPT-o1` (Chen et al., 2024a). This process builds upon the instruction tuning data and is designed to explicitly elicit and refine complex reasoning.

Given a dental image $\mathcal{I}$ and an question–answer pair (q^*, a^*) in instruction tuning data, the LLM is first prompted to analyze the question q^* and produce an initial reasoning path r_1 together with a tentative answer summary a_1:

$$(r_1, a_1) = \text{LLM}(q^*).$$

The generated answer a_1 is then compared with the reference answer a^* using the LLM as an evaluator. If a_1 is inconsistent with a^*, the LLM is prompted again with all previously generated information to continue the reasoning process and produce a refined reasoning step r_2 and answer a_2:

$$(r_t, a_t) = \text{LLM}(q^*, r_{1:t-1}, a_{1:t-1}), \text{until } a_t = a^*.$$

This iterative process continues until the generated answer matches the reference answer. All intermediate reasoning steps and answers $\{(r_1, a_1), (r_2, a_2), \dots\}$ are then concatenated to form a complete reasoning trajectory R.

Finally, conditioned on the dental image $\mathcal{I}$, the question q^*, and the aggregated reasoning trajectory R, the LLM generates a final summarized answer a:

$$a = \text{LLM}(\mathcal{I}, q^*, R).$$

The resulting data are formatted into a standardized CoT template of the form `<think>` R `</think><answer>` a `</answer>`, which trains a consistent response template and encourages reflective analysis, providing a cold start for the subsequent reinforcement learning stage.

C.2 Data Filtering

After completing all data construction steps, an additional quality control stage was applied using `GPT-5-mini` for secondary verification. Specifically, all Stage-I training samples constructed by `GPT-5` were re-checked, and samples that deviated from the original image content or reference labels

were filtered out. The retention rates after filtering are summarized in Table 5.

Data Type	PMC-Dental Caption	Opensource-Dental Classification	Opensource-Dental Detection
Image Captioning	96.63%	91.33%	93.80%
Instruction Tuning	96.85%	92.69%	95.49%
Total	46.7	63.2	67.1

Table 5: Retention rates after secondary verification with `GPT-5-mini`.

Overall, more than 90% of the generated samples were retained across all datasets and task types, indicating that the `GPT-5`-assisted data construction pipeline produced generally high-quality training data. In addition, datasets with richer and more detailed references tended to achieve higher retention rates. For example, `PMC-Dental-Caption`, which provides more descriptive image-level information, consistently showed the highest retention rates, whereas `Opensource-Dental-Classification`, which mainly contains disease labels with limited contextual information, yielded slightly lower retention.

C.3 Quality Assessment

As described in the methodology, the training dataset of DentalGPT was generated by `GPT-5` while referencing existing image labels or descriptions to minimize hallucinations and ensure professional domain knowledge injection. To validate the effectiveness of this approach, five evaluation dimensions were defined, and a randomly sampled set of 3,000 entries was assessed and compared against data obtained through direct `GPT-5` distillation. To ensure fairness, all comparisons were evaluated using `Gemini-2.5-Pro` as the judge.

Evaluation Setup Specifically, the dataset was evaluated across the following five dimensions: 1) *Description Completeness*: Whether all observable visual details in the image are thoroughly described, with particular attention to features that may contribute to dental diagnosis. 2) *Terminology Consistency*: Whether professional dental terminology is used correctly and consistently throughout the description. 3) *Content Safety*: Whether the content adheres to medical ethics and safety standards, avoiding sensitive, discriminatory, misleading, or inappropriate statements. 4) *Text–Image Consistency*: Whether the textual description is well written and accurately aligned with the corresponding image content. 5) *Knowledge Depth*: Whether the description demonstrates an appropriate level of dental knowledge.

`Gemini-2.5-Pro` was asked to score each dimension on a scale of 1 to 5, and the final dataset quality was reported using the average score across all evaluated samples.

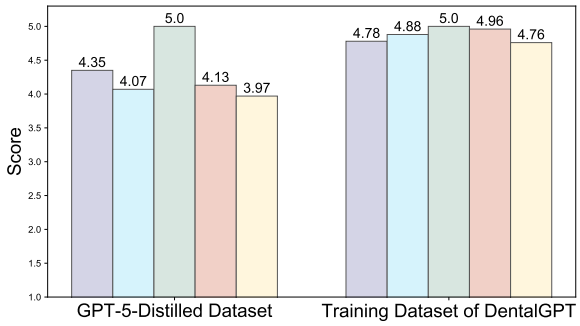


Figure 8: `Gemini-2.5-Pro`'s multi-dimensional evaluation of `GPT-5`–distilled data and the training dataset of DentalGPT (scores range from 0 to 5). For each dataset, the color blocks from left to right represent *Description Completeness*, *Terminology Consistency*, *Content Safety*, *Text–Image Consistency*, and *Knowledge Depth*.

Results Analysis Results are shown in Figure 8, our training dataset demonstrates clear advantages over the directly distilled version across multiple evaluation metrics. It can be observed that the data generated with label references shows the most significant improvements in *terminology consistency* and *knowledge depth*, and also achieves notable gains in the completeness of visual detail descriptions. Interestingly, `Gemini-2.5-Pro` assigns a perfect score for safety to both datasets, indicating that `GPT-5`–generated data perform exceptionally well in medical safety, avoiding harmful diagnostic suggestions and providing timely guidance to reduce potential risks.

In conclusion, the results indicate that our dataset, by leveraging annotations from public datasets and descriptions from academic literature, provides more comprehensive and more professional knowledge injection for the model. Such a foundation ensures substantial improvement in the performance of DentalGPT.

D Human Evaluation

To further assess the clinical value of DentalGPT, we conduct a human evaluation study to measure the degree to which licensed dentists prefer the model's outputs in realistic diagnostic scenarios.

Unlike automatic metrics, this evaluation focuses on whether the model's responses are aligned with professional clinical judgment and would be considered acceptable in practice.

Participants Four licensed dentists participated in the evaluation. All evaluators are professionally certified, with both the mean and median years of clinical practice being three years. As the evaluators are members of the same collaborative research project, compensation for participation was covered by the project's existing research funding.

Evaluation Data Evaluation images were randomly sampled from ten standard medical textbooks and three publicly available dental examination papers, resulting in a total of 80 dental images. These images cover a range of common dental conditions and realistic clinical scenarios. Each model was prompted to provide a detailed analysis of the image and a corresponding diagnostic conclusion.

Evaluation Protocol For each image, the response generated by DentalGPT was paired with that of a competing model. Dentists were asked to compare the two responses and indicate which one they preferred based on predefined criteria, including clinical accuracy, diagnostic soundness, and completeness of information. The prompt presented to the evaluators is shown in Figure 9. Each dentist evaluated all comparison pairs independently.

Result Aggregation For each model comparison, we aggregate the dentists' judgments and report the average proportions of *win*, *tie*, and *loss* outcomes. These statistics provide a quantitative measure of dentists' preferences and offer insight into the relative clinical acceptability of the model outputs.

The prompt for human evaluation

image

As a dental professional, please perform a rigorous clinical analysis of the provided image. Your response must prioritize **diagnostic accuracy**, the **absence of clinical errors**, and **comprehensive coverage** of all visible dental pathologies.

Please evaluate the image based on the following key dimensions: 1. **Clinical Correctness:** Ensure there are no theoretical or factual errors in the assessment. 2. **Diagnostic Accuracy:** Provide a precise diagnosis of the conditions shown (e.g., caries, periodontal status, restorations). 3. **Completeness:** Include all relevant visual details, such as tooth numbering (if applicable), location, and severity of the condition.

Use professional dentistry terminology while maintaining clinical relevance. Focus solely on the practical value of the diagnosis.

<caption>
{caption}
</caption>

<image context> (<image> represents the location of the image)
{image_context}
</image context>

Please output a detailed diagnostic analysis of the image only. Do not generate any content unrelated to the clinical task.

Figure 9: Prompt for human evaluation.

E Data Curation Prompts

All the prompts for data curation are listed below:

The prompt for Image Captioning

image

Please generate a professional, detailed, and high-quality description for the medical image I provide. The description should include as many detistry-related visual details as possible to ensure clinical readability, professionalism, and comprehensiveness, while also being understandable to the general public. Aim to make the description as rich and detailed as possible, providing extensive visual information. You may refer to the image's context to improve the accuracy and completeness of your description, but do not reveal that you referenced the context.

<caption>
{caption}
</caption>

<image context> (<image> represents the location of the image)
{image_context}
</image context>

Please output a detailed description of the image only. Do not generate any content unrelated to the task.

Figure 10: Prompt for generating observation-based dental image captions.

The prompt for Instruction Tuning: QA Generation

image

Please generate a dentistry-related question and a short corresponding ground truth answer about the medical image I provide. The question should assess the model's visual capabilities. Avoid being too specific—design the question so that it requires the model to look at the image to answer. The question should demand strong visual understanding as well as some knowledge of dentistry. You may refer to the provided image caption and contextual information to improve the quality of the question. However, **do not mention or reference the caption or context in the question itself—assume they are not available**.

<caption>
{caption}
</caption>

<image context> (<image> represents the location of the image)
{image_context}
</image context>

Please generate the question and the ground truth answer directly. Do not include any content unrelated to the task.

Figure 11: Prompt for generating dentistry-related question–answer pairs for instruction tuning.

The prompt for Instruction Tuning: Answer Generation

image

<question>
{caption}
</question>

<answer>
{ground_truth_answer}
</answer>

You are now required to look at the image I provide and answer the user's question about dentistry. Make sure your response directly addresses the user's query, follows instructions well, and is as detailed and rich as possible, with the style and quality characteristic of GPT-5.

You may refer to the image caption and contextual information I secretly provide to you in order to improve the accuracy and completeness of your answer. However, **do not mention or reference the caption or context in your response—assume they are not available**.

<caption>
{caption}
</caption>

<image context> (<image> represents the location of the image)
{image_context}
</image context>

Please generate the answer directly. Do not include any unrelated content.

Figure 12: Prompt for generating final answers during instruction tuning.

The prompt for answer verification

<Model Response>
{model_response}
</Model Response>

<Reference Answer>
{reference_answer}
</Reference Answer>

You are provided with a model-generated response (<Model Response>) and a reference answer (<Reference Answer>). Compare the model response with the reference answer and determine its correctness.

Task: Output `True` if the model response is correct, and `False` otherwise.

Please output only `True` or `False`, without any additional explanation.

Figure 13: Prompt for verifying the correctness of generated answers during iterative reasoning.

The prompt for Initial CoT Generation

<question>
{question}
</question>

Please respond to the above question <question> according to the provided image using the Chain of Thought (CoT) reasoning method. Your response should consist of multiple steps, each of which includes three types of actions: **"Inner Thinking"**, **"Final Conclusion"**, and **"Verification"**:

- **'Inner Thinking'**: This is the step where thinking is done. Note that multiple 'Inner Thinking' steps are required to describe thorough reasoning. Each step should first generate a brief title.
- **'Final Conclusion'**: At this stage, you summarize the correct reasoning from previous 'Inner Thinking' steps and provide the final answer. No title is required here.
- **'Verification'**: At this stage, you verify the conclusion from the "Final Conclusion" step. If the conclusion holds, end the process. If not, return to "Inner Thinking" for further reasoning. No title is required here.

The output format must strictly follow the JSON structure below:

“‘json "CoT": ["action": "Inner Thinking", "title": "...", "content": "...", ..., "action": "Final Conclusion", "content": "...", "action": "Verification", "content": "..."]

Figure 14: Prompt for initial construction of chain-of-thought reasoning.

The prompt for CoT Refinement with New Thinking

<question>
{question}
</question>

<previous reasoning>
{previous_reasoning}
<previous reasoning>

<response requirements>
Your response must include the following steps, each composed of three types of actions: **"Inner Thinking"**, **"Final Conclusion"**, and **"Verification"**:

1. **Inner Thinking**: Break down the reasoning process into multiple concise steps. Each step should start with a brief title to clarify its purpose.
2. **Final Conclusion**: Summarize the correct reasoning from all previous 'Inner Thinking' steps and provide the final answer. No title is needed for this section.
3. **Verification**: Verify the accuracy of the "Final Conclusion". If it holds, conclude the process. Otherwise, return to "Inner Thinking" for further refinement.
</response requirements>

<question> represents the image-based question to be answered, and <previous reasoning> contains your prior reasoning. Your task is to continue from the current 'Verification' step. I have manually reviewed the reasoning and determined that the **Final Conclusion** is false. Your 'Verification' results must align with mine. Proceed to refine the reasoning by exploring new approaches to solve this problem and construct a new Final Conclusion.

Output Format
Strictly follow the JSON structure below. You do not need to repeat your previous reasoning. Begin directly from the next 'Verification' stage.

“‘json "CoT": ["action": "Verification", "content": "...", "action": "Inner Thinking", "title": "...", "content": "...", ..., "action": "Final Conclusion", "content": "...", "action": "Verification", "content": "..."]

Figure 15: Prompt for refining CoT by exploring new reasoning strategies.

The prompt for CoT Refinement via Backtracking

<question>
{question}
</question>

<previous reasoning>
{previous_reasoning}

<previous reasoning>

<response requirements>
Your response must include the following steps, each composed of three types of actions: **"Inner Thinking"**, **"Final Conclusion"**, and **"Verification"**:

1. **Inner Thinking**: Break down the reasoning process into multiple concise steps. Each step should start with a brief title to clarify its purpose.
2. **Final Conclusion**: Summarize the correct reasoning from all previous 'Inner Thinking' steps and provide the final answer. No title is needed for this section.
3. **Verification**: Verify the accuracy of the "Final Conclusion". If it holds, conclude the process. Otherwise, return to "Inner Thinking" for further refinement.

</response requirements>

<question> represents the image-based question to be answered, and <previous reasoning> contains your prior reasoning. Your task is to continue from the current 'Verification' step. I have manually reviewed the reasoning and determined that the **Final Conclusion** is false. Your 'Verification' results must align with mine. Proceed to refine the reasoning using **backtracking** to revisit earlier points of reasoning and construct a new Final Conclusion.

Output Format
Strictly follow the JSON structure below. You do not need to repeat your previous reasoning. Begin directly from the next 'Verification' stage.

“‘json "CoT": ["action": "Verification", "content": "...", "action": "Inner Thinking", "title": "...", "content": "...", ..., "action": "Final Conclusion", "content": "...", "action": "Verification", "content": "..."]

Figure 16: Prompt for refining CoT via backtracking.

The prompt for CoT Refinement by Exploring New Image Regions

<question>
{question}
</question>

<previous reasoning>
{previous_reasoning}
<previous reasoning>

<response requirements>
Your response must include the following steps, each composed of three types of actions: **"Inner Thinking"**, **"Final Conclusion"**, and **"Verification"**:

1. **Inner Thinking**: Break down the reasoning process into multiple concise steps. Each step should start with a brief title to clarify its purpose.
2. **Final Conclusion**: Summarize the correct reasoning from all previous 'Inner Thinking' steps and provide the final answer. No title is needed for this section.
3. **Verification**: Verify the accuracy of the "Final Conclusion". If it holds, conclude the process. Otherwise, return to "Inner Thinking" for further refinement.
</response requirements>

<question> represents the image-based question to be answered, and <previous reasoning> contains your prior reasoning. Your task is to continue from the current 'Verification' step. I have manually reviewed the reasoning and determined that the **Final Conclusion** is false. Your 'Verification' results must align with mine. Proceed to refine the reasoning by focusing on new areas of the picture to solve this problem and construct a new Final Conclusion.

Output Format
Strictly follow the JSON structure below. You do not need to repeat your previous reasoning. Begin directly from the next 'Verification' stage.

```json "CoT": [ "action": "Verification", "content": "...", "action": "Inner Thinking", "title": "...", "content": "...", ..., "action": "Final Conclusion", "content": "...", "action": "Verification", "content": "..." ]
```

Figure 17: Prompt for refining CoT by exploring new visual regions.