OralGPT-Plus: Learning to Use Visual Tools via Reinforcement Learning for Panoramic X-ray Analysis

Supplementary Material

Appendix Contents

Appendix Contents

1. Related Work

1.1. VLMs for Dental Panoramic X-ray

Vision–language models (VLMs) have shown strong potential in oral radiology by unifying image understanding, dental knowledge, and diagnostic report generation. However, they face persistent challenges including limited domain-specific data, scarcity of fine-grained tooth-level annotations, factual reliability, and privacy constraints [7]. Recent dental VLMs [9] typically adopt a two-stage pipeline: (i) perform vision–language alignment or basic VQA-style modeling, and (ii) apply instruction tuning with rationales to enhance open-ended descriptions and clinical coherence [18]. Parallel to these efforts, several domain-specific datasets and benchmarks have emerged. OralGPT [5] introduces a five-dimension evaluation showing that both general and medical VLMs exhibit large performance gaps on tooth-level and region-grounded diagnostic tasks, but that in-domain instruction tuning can markedly improve accuracy. DentalBench [22] further provides the first comprehensive bilingual benchmark and corpus for dental LLM evaluation, demonstrating that domain adaptation is crucial for performing specialized tasks such as intervention classification or pathology interpretation. Methodologically, these works often rely on structured cues—such as FDI tooth indices, treatment records (filling, crown, RCT, implant), and similarity-based pseudo-labeling—to compensate for supervision shortages and improve interpretability. Yet despite superior textual expressiveness, existing models still rely on single-pass inference and lack explicit representations of the multi-step inspection and contralateral comparison that dentists routinely perform. Unlike prior single-pass dental VLMs, OralGPT-Plus introduces the first agentic panoramic dental VLM capable of iterative diagnostic reasoning. Through curated multi-turn trajectories (DentalProbe), explicit diagnostic tools (“Zoom-In” + “Mirror-In”), and reinspection-driven reinforcement learning, OralGPT-Plus aligns directly with real clinical workflows rather than relying solely on static image–text alignment.

1.2. Thinking with Images in VLMs

Recent advances in multimodal reasoning have introduced the paradigm of “thinking with images,” where VLMs dynamically interact with images through tools such as zooming, cropping, and region selection [3, 8, 11, 16, 17, 21]. This allows models to refine hypotheses and perform multi-step visual analysis. However, existing approaches lack medical priors, do not leverage expert multi-step trajectories, and may exhibit uncontrolled exploration, making them unsuitable for clinical diagnosis. Panoramic dental radiographs naturally fit this paradigm, as dentists routinely zoom into subtle regions and compare symmetric structures. Yet prior work has not adapted thinking-with-images to panoramic diagnosis. We address this gap by constructing DentalProbe for expert-guided trajectories, introducing “Mirror-In” for symmetry-aware inspection, and developing a reinspection-driven RL framework that aligns exploration with diagnostic reliability.

2. Details of DentalProbe Curation

2.1. Data Statistics Details

As shown in Tab. 2, we integrate four publicly available panoramic radiograph datasets and select part of suitable

Table 1. Categories of radiograph findings and abnormal conditions.

Some categories of radiograph findings and abnormal conditions		
Apical Periodontitis	Periapical Abscess	Periapical Granuloma
Apical Sclerosing Osteitis	Apical Rarefying and Sclerosing Osteitis	Focal Osteomyelitis
Pericoronitis	Periodontal Disease / Bone Loss	Endo-Perio Combination Lesion
Implantitis / Peri-implantitis	Mucous Retention Pseudocyst	Osteoarthritis of the Condyle
Tonsilloliths	Calcified Carotid Atheromatous Plaque	Antrolith
Foreign Body Reaction	Oroantral Communication	Trauma from Occlusion
Dentigerous Cyst	Odontogenic Keratocyst (OKC)	Radicular Cyst
Incisive Canal Cyst	Ameloblastoma	Odontoma
Osteoma	Malignancy (Suspected)	Impacted Tooth
Supernumerary Tooth / Mesiodens	Retained Deciduous Tooth	Hyperplastic Follicle
Dilacerated Tooth	Microdont	Hypercementosis
Dense Bone Island / Idiopathic Osteosclerosis	Condylar Hyperplasia	Condylar Hypoplasia
Atypical Condylar Morphology	Large Marrow Defect	Stafne Defect
Calcification of Stylohyoid Ligament	Apical Cemento-Osseous Dysplasia	Florid Osseous Dysplasia
Focal Cemento-Osseous Dysplasia	Tooth Fracture / Vertical Root Fracture	Jaw / Condylar Fracture
Osteopenia / Osteoporosis	Remnant Root Fragment	Root Resorption
Disuse Alveolar Atrophy	Ankylosis	Displacement of Fracture Fragment
Prosthetic Restoration	Orthodontic Device	Surgical Device
Implant	Bone-based Findings	Carious Lesion
Furcation Lesion		

images: the Tufts Dental Database (380 images) [12], the Children's Panoramic Dataset (69 high-quality pediatric cases) [19], the fully annotated DENTEX Challenge dataset (678 images) [4], and the Diagnostics TEAM dataset (1435 images) [10]. Together, these sources contribute a total of 2,562 clinically annotated panoramic radiographs covering multiple demographic groups, including both children and adults, and originating from geographically diverse regions such as the United States, China, Switzerland, Turkey, and Romania. Across these datasets, more than 50 dental conditions are labeled, including caries, periapical lesions, impacted teeth, periodontal bone loss, prosthetic restorations, orthodontic appliances, and rare developmental anomalies (the details of diseases list are shown in Tab. 1). DENTEX and Tufts provide tooth-level enumeration that follows the FDI two-digit notation system, while the pediatric dataset contributes mixed-dentition cases that reflect developmental-stage variability. This diverse and globally distributed composition substantially enhances the representativeness, robustness, and clinical generalizability of the final DentalProbe dataset and MMOral-X benchmark.

2.2. Trajectory Construction Details.

In this section, we introduce the details of how we build the trajectories of DentalProbe dataset.

To construct high-quality diagnostic trajectories, we design a multi-stage, multi-agent generation process that emulates the reasoning workflow of dentists on panoramic X-rays.

Step 1: Tooth Detection and Lesion Alignment. For each OPG sample, a tooth-number detection expert model identifies all tooth bounding boxes and aligns them with the corresponding ground-truth radiograph findings annotations. Findings are categorized into three types shown in Tab. 4: (a) obvious findings, (b) subtle findings, and (c) bone-based findings. This taxonomy guides the subsequent proposal generation and reasoning flow.

Step 2: Proposal Generation. For subtle and bone-based findings, we apply k-means clustering to the detected tooth-level bounding boxes to generate multiple region proposals. Each proposal represents a potentially ambiguous or hard-to-diagnose area requiring focused analysis in later reasoning steps.

Table 2. Summary of Integrated Public Dental Datasets

Dataset	Images Used	Annotation Scope	Region
Tufts [12]	380	Tooth & lesion labels; 5-level taxonomy	USA
Children [19]	69	Pediatric segmentation & disease labels	China
DENTEX [4]	678	Quadrant, enumeration, diagnosis (4 types)	Switzerland/Turkey/USA
Diagnostics TEAM [10]	1435	Detection of 12 clinical conditions	Romania
Total	**2562**	Fine-grained abnormality annotations	Multi-region

Step 3: Multi-Agent Trajectory Synthesis We then construct multi-turn reasoning trajectories under a rule-based template:

- **Round 1:** The initial query Q_1 focuses on global inspection. The model outputs A_1 emphasizing obvious findings and triggers a "Zoom-In" tool call to the detected region.
- **Round 2 and onward:** The zoomed-in result becomes the next-round input Q_{i+1}, where the query template is extended with contextual cues (*"After the above Action X..."*). The model continues reasoning and generating the corresponding answer A_{i+1}.
- **Tool Selection with Multi-Agent Collaboration:** For each zoomed-in image, answer agent performs diagnostic Q&A, while judge agent evaluates the answerability. If the case can be correctly answered, the proposal is retained under "Zoom-In"; otherwise, the system invokes "Mirror-In" to create a symmetric comparison view. visual describe agent then produces detailed visual descriptions for each step.

We employ GPT-5 [11] to perform different roles of this multi-agent process, the prompts are shown in Fig. 14, Fig. 15, Fig. 16, Fig. 17. This iterative process yields a complete multi-turn diagnostic trajectory incorporating both local inspection and symmetry-aware comparison.

Step 4: Linguistic Enhancement and Visual Alignment. For each Q/A pair, we refine textual reasoning by integrating two complementary sources: (i) standard radiographic descriptions retrieved from Wikipedia according to the ground-truth disease category, and (ii) visual features extracted by agent 3. The merged content is rewritten into coherent diagnostic narratives, ensuring consistency between image evidence and textual explanation. For each category, the radiographic descriptions we use can be seen in Tab. 3.

Through this multi-agent, multi-round synthesis pipeline, we obtain high-quality trajectories that encapsulate clinical reasoning patterns, structured tool usage, and visually grounded explanations—serving as the foundation for Dentist-like Instruction Tuning.

2.3. Teeth ID Detection.

For teeth id detection, we use the expert model from [5] trained with 2796 samples from 3 open-sourced datasets as shown in Tab. 5, which labeled with 1 to 32 tooth numbering following the FDI tooth numbering system. The teeth detection ability is powerful and robust.

2.4. Dental-Aware Tool Design Details

Conventional agentic visual reasoning models commonly rely on the "Zoom-In" operation to enlarge suspicious regions for detailed observation. However, such a purely local enhancement often neglects the inherent anatomical symmetry of the oral cavity. In clinical practice, dentists routinely perform "contralateral comparison" to analyze the corresponding tooth or region on the opposite side of the jaw—to evaluate whether a subtle finding (e.g., shadow, opacity, or margin irregularity) reflects a true lesion or merely a radiographic artifact. This diagnostic habit leverages the structural bilateral symmetry of the dentition and surrounding bone, providing a natural reference baseline for uncertain cases.

Inspired by this clinical rationale, we transform oral symmetry from an implicit prior into an explicit and callable diagnostic action through a novel tool, 'Mirror-In'. The "Mirror-In" operation allows the model, after localizing a potential lesion, to automatically retrieve the horizontally symmetric region across the midline, thereby constructing a paired comparison view that emulates human inspection behavior. Formally, let $I(x, y)$ denote the panoramic X-ray image of width W. For a selected region $[x_1, x_2] \times [y_1, y_2]$, the mirrored counterpart is defined as:

$$\begin{aligned} I_{\text{mirror}}(x, y) &= I(W - x,\, y), \\ &(x, y) \in [x_1, x_2] \times [y_1, y_2]. \end{aligned} \tag{1}$$

This operation reflects pixel coordinates along the vertical midline and returns a horizontally flipped counterpart of the selected patch, forming a structured dual-view pair $(I_{\text{crop}}, I_{\text{mirror}})$ for subsequent reasoning.

Compared with the conventional "Zoom-in" that focuses solely on intra-patch details, the "Mirror-In" tool introduces an inter-patch relational perspective. It enables the model to

Table 3. Categories and their radiographic descriptions retrieved from wiki and google webs.

Category	Radiographic Description
Abnormal tooth development	On panoramic radiographs, developmental abnormalities may involve tooth number, size, shape, or internal structure. Common findings include supernumerary or missing teeth, macrodontia, microdontia, dilaceration, dens in dente, ankylosis indicated by loss of periodontal ligament space, impacted or unerupted teeth, and enlarged pulp chambers such as taurodontism.
Caries	Caries typically presents as a radiolucent area within the enamel or dentin. Shallow lesions affect the enamel or outer dentin, moderate lesions reach the mid-dentin, and deep lesions extend toward the pulp. Early lesions may be subtle due to overlapping anatomy and reduced resolution in panoramic imaging.
Deep pits and fissures	Deep occlusal pits and fissures are often poorly visualized on panoramic radiographs because anatomical overlap hides early decay. Only advanced lesions penetrating dentin may appear as radiolucent defects, while early changes remain masked.
Periapical periodontitis	Appears as a radiolucent area surrounding the root apex or as widening of the periodontal ligament space. Early lesions may only present slight thickening or loss of lamina dura and can be difficult to detect due to superimposed anatomical structures.
Pulpitis	Chronic pulpitis may show widening of the periodontal ligament space or discontinuity of the lamina dura. Advanced pulpal inflammation may lead to periapical radiolucency or areas of increased bone density (condensing osteitis).
Impacted tooth	An impacted tooth appears as an unerupted or malpositioned tooth within the jawbone. Its outline may be partially obscured or distorted by overlapping structures, resulting in ghost images, blurry contours, or uncertain boundaries.
Periapical lesion	A periapical lesion appears as an irregular radiolucent area at the root apex. Small lesions may not be clearly visible due to the lower resolution of panoramic radiographs, often necessitating periapical imaging or CBCT.
Carious lesion	Carious lesions appear as radiolucent zones reflecting demineralization. Proximal caries often display a triangular radiolucency with the base at the surface, whereas early or mild lesions may appear as subtle lines, dots, or shadows that can be confused with artifacts.
Apical periodontitis	Presents as a periapical radiolucency or widened periodontal ligament space caused by inflammatory bone loss at the root apex. Detection may be limited by superimposed anatomical structures, and absence on panoramic imaging does not exclude disease.
Furcation lesion	A furcation lesion appears as a radiolucent region between the roots of multi-rooted teeth, representing bone loss or periodontal ligament widening. Early-stage involvement may be poorly visualized on panoramic radiographs.
Root resorption	Root resorption shows as radiolucent defects along the root surface or within the root structure. External resorption produces irregular root outlines, while internal resorption appears as a smooth, symmetrical enlargement of the pulp chamber or canal.
Root fragment	A retained root fragment appears as a distinct radiopaque fragment separated from the main root, often irregular in shape and embedded in the alveolar bone.
Bone resorption	Bone resorption appears as thinning or loss of alveolar bone height, reduced cortical density, and less-defined cortical outlines. In some cases, an onionskin-like appearance may be present.

evaluate whether a suspected finding deviates from its symmetric counterpart, effectively providing a self-calibrated “reference control.” This mechanism substantially mitigates diagnostic uncertainty arising from local exposure imbalance, motion blur, or anatomical variation. For instance, small carious lesions, enamel defects, or periapi-

Table 4. Three categories of radiographic findings used in our constructed multi-turn trajectories of DentalProbe dataset. Obvious findings appear only in Round 1, bone-based findings and subtle findings appear in Round 2 and onward, Due to the large number of subtle findings, we display only a subset of this type.

Category	Finding Type	Appearing Round(s)
Obvious findings	Prosthetic restoration	Round 1
	Orthodontic device	Round 1
	Surgical device	Round 1
	Implant	Round 1
	Impacted tooth	Round 1
Bone-based findings	Bone resorption	Round 2+
Subtle findings	Carious lesion	Round 2+
	Apical periodontitis	Round 2+
	Furcation lesion	Round 2+
	Root resorption	Round 2+
	Root fragment	Round 2+

Table 5. The details of the teeth ID detection model we use.

Dataset Source	Task Type	Category Space	Categories	Training Set	mAP	AP50
[1, 4, 12]	Object Detection	1 to 32 tooth numbering following the FDI tooth numbering system	32	2798	66.1	98.9

cal radiolucencies can be better verified by comparing with the healthy contralateral region. From a modeling standpoint, "Mirror-In" enriches the agent's observation space with cross-quadrant contrastive cues, promoting relational consistency and spatial awareness beyond the local receptive field.

Clinically, the "Mirror-In" design resonates with the dentist's habitual reasoning process—"see, compare, and conclude"—while computationally it offers a structured way to embed symmetry priors into visual reasoning loops. By enabling symmetry-aware contrastive inspection, our agent achieves more stable judgments for subtle or low-contrast abnormalities and demonstrates stronger generalization under radiographic noise and patient-specific variation.

Table 6. Stability verification of using LLMs as judges: Standard deviation and coefficient of variation (CV) are reported across four VLMs from five repeated evaluations.

Model	Mean	StdDev	CV %
GPT-5 [11]	32.88	0.18	0.54
Gemini-2.5-Flash [15]	25.60	0.24	0.96
GLM-4.5v [6]	15.70	0.33	2.16
Qwen2.5-VL-7B-Instruct [2]	7.20	0.34	4.71

3. Stability and Consistency for Evaluation

In this section, we present the systemmatic analysis of the evaluation metric for MMOral-X. The analysis can be seen in 6 Since using LLMs as judges inevitably introduces randomness, even with the temperature hyperparameter set to 0, we conduct multiple repeated experiments to verify the stability of LLMs as judges. Specifically, we evaluate the prediction results of GPT-5 [11], Gemini-2.5-Flash [15], GLM-4.5v [6], and Qwen2.5-VL-7B-Instruct [2] on the MMOral-X benchmark using GPT-5-mini [11] with the same prompt five times. The obtained mean, standard deviation, and coefficient of variation (CV) of the metric "overall" are shown in Table 6. For proprietary models, medical-specific models, and general-purpose LVLMs, the standard deviation of the metric "overall" is no more than 0.212 when evaluated 5 times using GPT-5-mini with our designed few-shot prompt. Specifically, for the prediction results of GPT-5, the standard deviation of the scores is 0.179, while for Qwen3-VL-8B, the standard deviation is as low as 0.039. Meanwhile, CV (Coefficient of Variation), as a standardized measure of dispersion of a probability distribution, can be used to assess the stability of scores across multiple experiments. The CV values for the prediction results of these four models, after being scored 5 times, are all around 0.5%, which demonstrates the evaluation stability of using LLMs as evaluators.

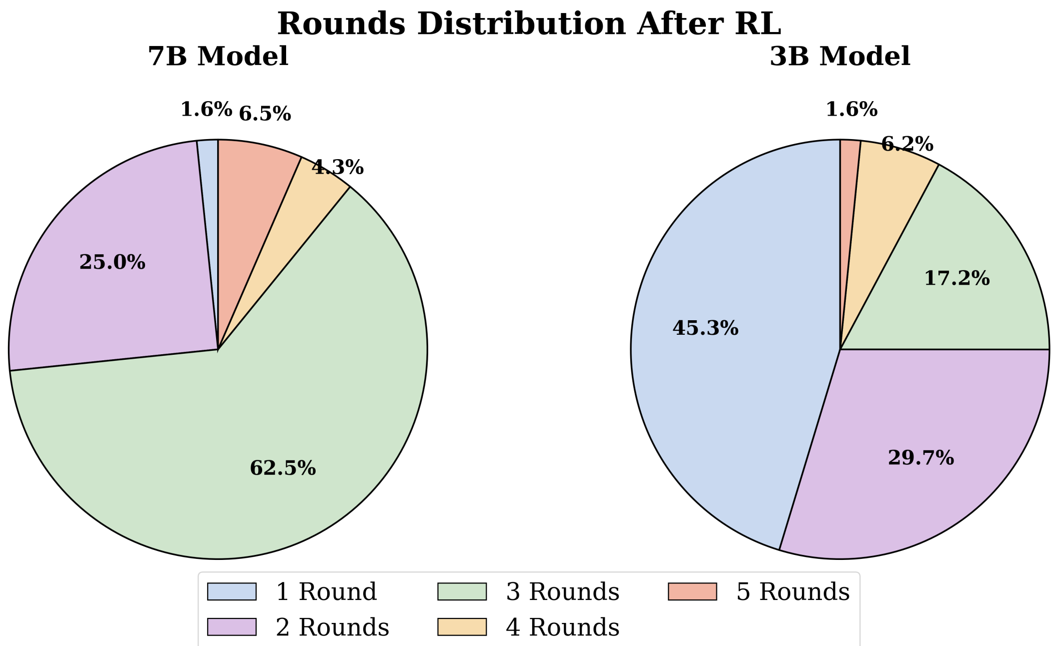


Figure 1. **Rounds distribution after RL of OralGPT-Plus 7B and 3B.**

4. Training Details

Table 7. Hyperparameters for training Qwen2.5-VL (7B and 3B) models in dentist-like instruction tuning.

Hyperparameter	**Value**
Epoch	3
LoRA Rank	8
LoRA α	16
LoRA Dropout	0.1
LoRA Target	all
GPU	4 $\times$ NVIDIA A800
Batch Size	16
Gradient Accumulation Steps	8
Warmup Ratio	0.1
Learning Rate	1e-4
Learning Rate Scheduler	Cosine
Unfreeze Vision Tower	True

4.1. Hyperparameters of Training.

The hyperparameters of dentist-like instruction tuning and Reinspection-driven reinforcement learning are shown in Tab. 7 and Tab. 8. For instruction tuning, we follow the basic setting of LLaMA-Factory [20]. For reinforcement learning, we follow the setting of Mini-o3 [8] using Verl framework [14].

4.2. Rubrics-based Reward

For rubrics-based reward, the prompt we use for can be seen at Fig. 12 and Fig. 13.

4.3. Conditioned Diagnostic-Driven Reward

Following the confidence-gated curiosity mechanism in PixelReasoner [16], we design a conditioned diagnostic-driven reward that activates visual reinspection only when the textual diagnosis is already reliable but potentially incomplete. For a query x and a generated trajectory τ, let $R_{\text{rubrics}}(\tau) \in [0, 1]$ denote the rubric-based diagnostic reliability, and let $n_{\text{tool}}(\tau)$ be the number of visual tool invocations in τ. We define the exploration saturation level for query x as the expected tool-usage intensity under the current policy

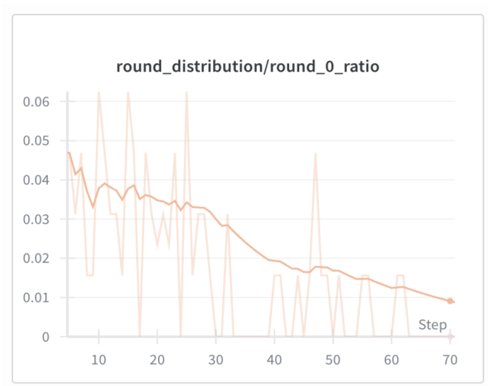
$$\mathrm{Cu}(x) \doteq \mathbb{E}_{\tau \sim \pi_\theta(\cdot|x)} \big[n_{\text{tool}}(\tau) \big] \approx \frac{1}{T} \sum_{t=1}^{T} n_{\text{tool}}(\tau_t), \quad (2)$$

where $\{\tau_t\}_{t=1}^{T}$ are recent rollouts for the same query x. We further define the binary indicator $\mathbb{1}_{\text{tool}}(\tau) = \mathbb{I}\big(n_{\text{tool}}(\tau) > 0\big)$, which records whether any visual tool is used in τ.

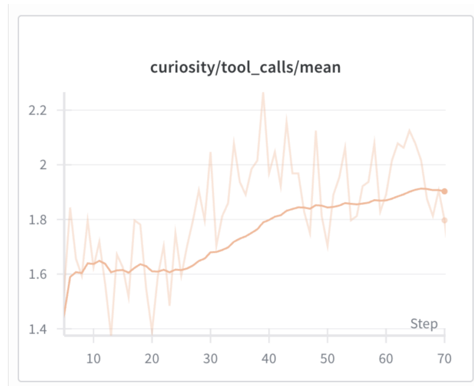
Applying a Lagrangian-style confidence gating as in PixelReasoner [16], we obtain the intrinsic curiosity bonus

$$R_{\text{diag}}(\tau) = \alpha \, \mathbb{I}\big(R_{\text{rubrics}}(\tau) > \eta\big) \, \big(H - \mathrm{Cu}(x)\big)_+ \mathbb{1}_{\text{tool}}(\tau), \quad (3)$$

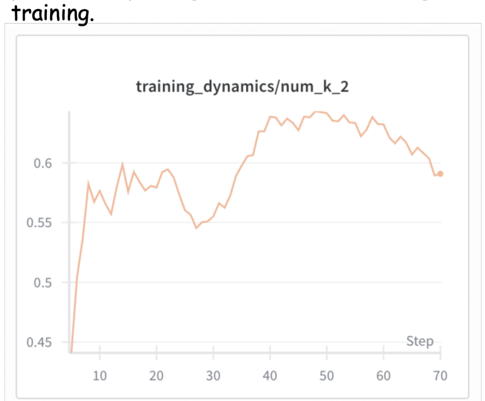
where $(\cdot)_+ = \max(\cdot, 0)$ ensures that the bonus decays to zero as $\mathrm{Cu}(x)$ approaches the saturation level H. This confidence-conditioned formulation, directly following the spirit of PixelReasoner, encourages clinically meaningful reinspection only when the diagnosis is sufficiently trustworthy and the visual exploration budget for query x is not yet exhausted.



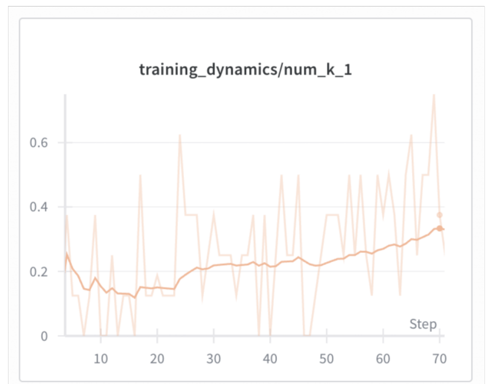
(a) Change in the proportion of samples that produce only a single-round answer during training.



(b) Changes in the number of tool calls during training.



(c) Changes in the proportion of tool-using samples that contain both correct and incorrect predictions during training.



(d) Change in the proportion of tool-using samples whose responses are entirely correct during training.

Figure 2. Evolution of key training dynamics throughout reinspection-driven reinforcement learning.

4.4. RL Optimization Objective

Here we provide a detailed description of the optimization objective used in our reinspection-driven reinforcement learning procedure. We optimize the policy using the GRPO [13] objective with clipped importance ratios, following a PPO-style constrained update strategy that stabilizes reinforcement learning on long-horizon, multi-turn reasoning tasks. To avoid overloading the action notation, we use $\mathbb{A}_i$ to denote the standardized advantage associated with the i-th sampled response. The optimization objective is defined as:

$$\mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}_{q\sim\mathcal{D},\, o_i\sim\pi_{\theta_{\text{old}}}}\left[\frac{1}{G}\sum_{i=1}^{G}\min\bigl(r_i\,\mathbb{A}_i,\ \hat{r}_i\,\mathbb{A}_i\bigr)\right], \quad (4)$$

$$r_i = \frac{\pi_\theta(o_i\,|\,q)}{\pi_{\theta_{\text{old}}}(o_i\,|\,q)}, \quad \hat{r}_i = \text{clip}(r_i,\, 1-\epsilon,\, 1+\epsilon) \quad (5)$$

$$\mathbb{A}_i = \frac{r_i - \bar{r}}{s_r}, \qquad \bar{r} = \tfrac{1}{G}\sum_{j=1}^{G} r_j, \quad (6)$$

$$s_r = \sqrt{\tfrac{1}{G}\sum_{j=1}^{G}(r_j - \bar{r})^2}. \quad (7)$$

In this formulation, each query q is sampled from the training distribution $\mathcal{D}$, and for every query the old policy $\pi_{\theta_{\text{old}}}$ generates G candidate outputs $\{o_i\}$, which are then evaluated to produce graded rewards. The term r_i is the importance ratio that measures how the new policy π_θ changes the likelihood of producing the same output compared with the old policy, while the clipped version $\hat{r}_i$ constrains the update within the trust region $[1-\epsilon, 1+\epsilon]$ to avoid excessively large policy shifts. The advantage $\mathbb{A}_i$ is computed by standardizing the set of ratios $\{r_i\}$ within the same query, making $\mathbb{A}_i$ a relative measure of how much better the i-th sample performs compared with other candidates for that query. This intra-query normalization stabilizes optimization and prevents the model from overfitting to a small number of high-reward trajectories. The GRPO objective then takes the minimum of the unclipped and clipped surrogate terms, ensuring that beneficial updates are encouraged while overly aggressive updates are suppressed. Through this mechanism, the model can reliably propagate graded, clinically grounded rewards into stable advantages, enabling effective learning of multi-step, symmetry-aware reasoning behaviors without collapsing into degenerate tool-use patterns.

Table 8. Key hyperparameters for resinspection-driven RL training in OralGPT-Plus.

Hyperparameter	**Value**
Data & Prompt Settings	
System Prompt	tool_crop_mirror
Train Batch Size	8
Max Prompt Length	8192
Max Response Length	8192
Image Key	images
Answer Key	solution
Max Pixels / Min Pixels	2M / 40k
Tool Call Type	crop_mirror
Reward & Algorithm	
Advantage Estimator	GRPO
Accuracy Reward Weight	1.0
Format Reward Weight	0
Tool Call Penalty	0
KL Coefficient	0.001
Actor & RL Training	
Learning Rate	1e-6
PPO Mini-batch Size	4
PPO Micro-batch / GPU	1
Clip Ratio (high / low)	0.3 / 0.2
Use KL Loss	False
Entropy Coefficient	0.000
Gradient Checkpointing	True
Training Runtime	
Rollout Batch Size	8
Max Tokens Per Batch	32768
GPU Memory Utilization	0.6
GPU	4

5. Additional Analysis

5.1. Analysis on rounds distribution

After reinforcement learning, the two models exhibit distinctly different round distributions, highlighting the impact of model capacity on multi-turn diagnostic behavior. As shown in Fig. 1, the 7B model predominantly adopts longer trajectories: 62.5% of its cases use three rounds and 25.0% use two rounds, while only 1.6% terminate after a single round. This pattern indicates that the larger model consistently conducts deeper reinspection and maintains stable tool-based reasoning throughout the diagnostic process. In contrast, the 3B model shows a much flatter distribution. Nearly 45.3% of its outputs stop after a single round, and only 17.2% reach three rounds. The early-termination behavior suggests that the smaller model struggles to sustain multi-step examination and often fails to trigger additional clinically meaningful inspections.

These observations are fully consistent with our earlier findings that model capacity determines the effectiveness of reinforcement learning. The 7B model reliably develops a tool-driven diagnostic workflow and benefits substantially from RL, whereas the 3B model gains much less due to its limited ability to support stable, multi-round tool usage.

5.2. Qualitative Analysis of Tool-Usage Behaviors

During reinspection-driven reinforcement learning, our model exhibits clear behavioral transitions that illustrate the emergence of a structured, dentist-like diagnostic workflow. Figure 2 summarizes four complementary and key trends reflecting how tool usage, multi-round reasoning, and correctness evolve throughout training.

First, the proportion of single-round responses steadily decreases, indicating that the model progressively abandons simplistic one-shot predictions in favor of multi-step inspection and verification. At the same time, the average number of tool calls consistently increases, suggesting a growing reliance on zooming, mirroring, and other visual operations as the model learns to actively explore ambiguous regions rather than passively interpreting the raw image.

The fraction of tool-using samples that contain both correct and incorrect predictions initially rises and later stabilizes, reflecting a natural exploration–to–convergence trajectory. Early in training, the model experiments with diverse tool-invoked reasoning patterns; as learning progresses, this exploratory variability gradually becomes more structured. Finally, the proportion of tool-using samples whose outputs are entirely correct grows steadily, demonstrating that the model not only uses tools more frequently but also uses them more effectively. This trend confirms that the conditioned diagnostic-driven reward successfully guides the agent toward clinically meaningful reinspection behaviors and more reliable diagnostic conclusions.

Overall, these dynamics reveal a consistent shift from passive image understanding to structured, tool-mediated multi-round reasoning, which is highly consistent with the observed performance improvements.

5.3. Scalability Analysis

The architectural design of OralGPT-Plus naturally supports scalability across model capacity, tool kinds, and reinforcement learning complexity. Larger models demonstrate stronger gains from the proposed training pipeline, as capacity directly enhances the stability of multi-turn reasoning and the reliability of tool usage. This indicates that the framework can continue to improve when scaled to models with higher parameter counts, where richer intermediate representations and more consistent inspection policies are likely to emerge.

In addition, both the tool interface and the reinforcement learning framework are constructed in a modular manner. The current thought–action–observation loop remains compatible with additional inspection tools such as multi-scale visual operators, contrast manipulation, or region highlighting, without requiring changes to the core architecture. The hybrid reward structure also supports further extension, since each reward term is defined independently and can incorporate new clinical objectives as task complexity increases. Finally, although OralGPT-Plus is developed for panoramic radiographs, the underlying reasoning paradigm generalizes to other clinical imaging domains that require localized examination and iterative verification, such as mammography, fundus imaging, CT interpretations, or ultrasound analysis. These properties collectively show that the system is well suited for scaling toward broader and more advanced medical agent applications.

5.4. Failure Analysis

Failure cases mainly arise in highly complex radiographs where multiple subtle, overlapping, or low-contrast lesions appear simultaneously. In such settings, even well-timed tool usage cannot fully resolve visual ambiguity: magnified views may emphasize noise or anatomical variations, and symmetry-based comparisons may surface misleading cues. As a result, the model may over-interpret minor artifacts, omit clinically important findings, or generate confident yet incorrect explanations. These observations indicate that when a single image contains dense and visually entangled abnormalities, tool-augmented reasoning may still be insufficient and can occasionally amplify hallucinations rather than suppress them.

6. Additional Cases

Proposal Cases. Additional proposals we clustered are shown in Fig. 3, Fig. 4, Fig. 5 and Fig. 6.

DentalProbe Cases. Additional trajectories cases of DentalProbe are shown in Fig. 7, Fig. 8 and Fig. 9

MMOral-X Cases. Additional MMOral-X cases and comparison between the response of OralGPT-Plus and other models are shown in Fig. 10 and Fig. 11.

References

[1] Dentistry computer vision project, 2023. 5

[2] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-vl technical report. *ArXiv*, abs/2502.13923, 2025. 5

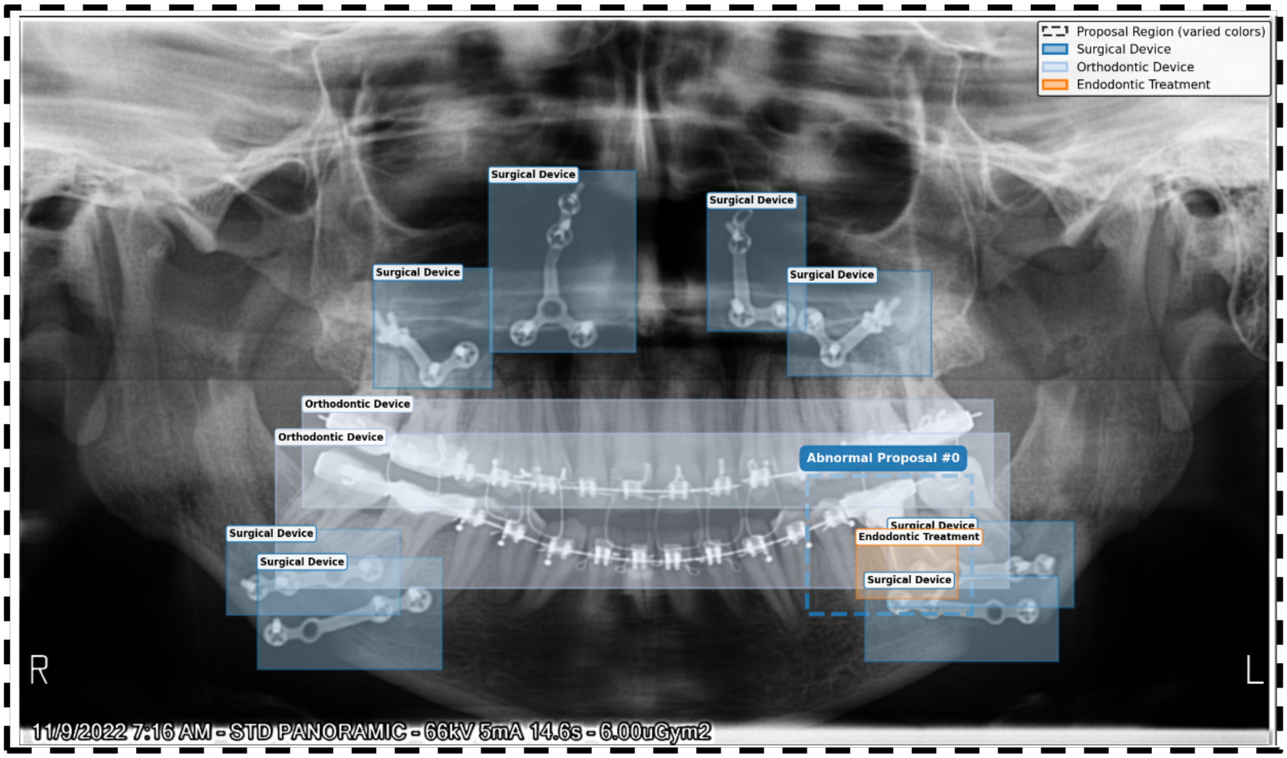
[3] Xinyu Geng, Peng Xia, Zhen Zhang, Xinyu Wang, Qiuchen Wang, Ruixue Ding, Chenxi Wang, Jialong Wu, Yida Zhao, Kuan Li, et al. Webwatcher: Breaking new frontier of vision-language deep research agent. *arXiv preprint arXiv:2508.05748*, 2025. 1

[4] Ibrahim Ethem Hamamci, Sezgin Er, Enis Simsar, Atif Emre Yuksel, Sadullah Gultekin, Serife Damla Ozdemir, Kaiyuan Yang, Hongwei Bran Li, Sarthak Pati, Bernd Stadlinger, et al. Dentex: An abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic x-rays. *arXiv preprint arXiv:2305.19112*, 2023. 2, 3, 5

[5] 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. Towards better dental ai: A multimodal benchmark and instruction dataset for panoramic x-ray analysis. *The Thirty-Ninth Annual Conference on Neural Information Processing Systems*, 2025. 1, 3

[6] GLM-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, Aohan Zeng, Baoxu Wang, Boyan Shi, Changyu Pang, Chenhui Zhang, Da Yin, Fan Yang, Guoqing Chen, Jiazheng Xu, Jiali Chen, Jing Chen, Jinhao Chen, Jinghao Lin, Jinjiang Wang, Junjie Chen, Leqi Lei, Letian Gong, Leyi Pan, Mingzhi Zhang, Qinkai Zheng, Shengchao Yang, Shilong Zhong, Shiyu Huang, Shuyuan Zhao, Siyan Xue, Shangqing Tu, Shengbiao Meng, Tianshu Zhang, Tian-Yuan Luo, Tianxiang Hao, Wenkai Li, Wei Jia, Xinpeng Lyu, Xuancheng Huang, Yanling Wang, Ya-Qi Xue, Yanfeng Wang, Yifan An, Yifan Du, Yi Shi, Yiheng Huang, Yilin Niu, Yuan Wang, Yuanchang Yue, Yuchen Li, Yutao Zhang, Yuxuan Zhang, Zhanxiao Du, Zhenyu Hou, Zhao Xue, Zhengxiao Du, Zihan Wang, Peng Zhang, De-Huan Liu, Bin Xu, Juanzi Li, Minlie Huang, Yuxiao Dong, and Jie Tang. Glm-4.5v and glm-4.1v-thinking: Towards versatile multimodal reasoning with scalable reinforcement learning. 2025. 5

[7] Hanyao Huang, Ou Zheng, Dongdong Wang, Jiayi Yin, Zijin Wang, Shengxuan Ding, Heng Yin, Chuan Xu, Renjie Yang, Qian Zheng, et al. Chatgpt for shaping the future of



Description:
Because Surgical Device and Orthodontic Device are classified as obvious findings, they are excluded from the proposal clustering process.
Question:
What irregular findings are shown in this panoramic dental X-ray?
Answer:
Regions [0.28, 0.35, 0.38, 0.51], [0.37, 0.21, 0.49, 0.46], [0.55, 0.25, 0.63, 0.43], [0.61, 0.35, 0.73, 0.49], [0.17, 0.70, 0.30, 0.82], [0.19, 0.74, 0.34, 0.90], [0.69, 0.69, 0.84, 0.81], [0.67, 0.77, 0.83, 0.89] contain surgical devices. Regions [0.23, 0.53, 0.77, 0.67] and [0.20, 0.57, 0.79, 0.78] contain orthodontic devices. Endodontic treatment is present at tooth 36 with a periapical radiolucency suggestive of periapical periodontitis or a healing lesion.

Figure 3. Case of clustered proposal.

dentistry: the potential of multi-modal large language model. *International Journal of Oral Science*, 15(1):29, 2023. 1

[8] Xin Lai, Junyi Li, Wei Li, Tao Liu, Tianjian Li, and Hengshuang Zhao. Mini-o3: Scaling up reasoning patterns and interaction turns for visual search. *arXiv preprint arXiv:2509.07969*, 2025. 1, 6

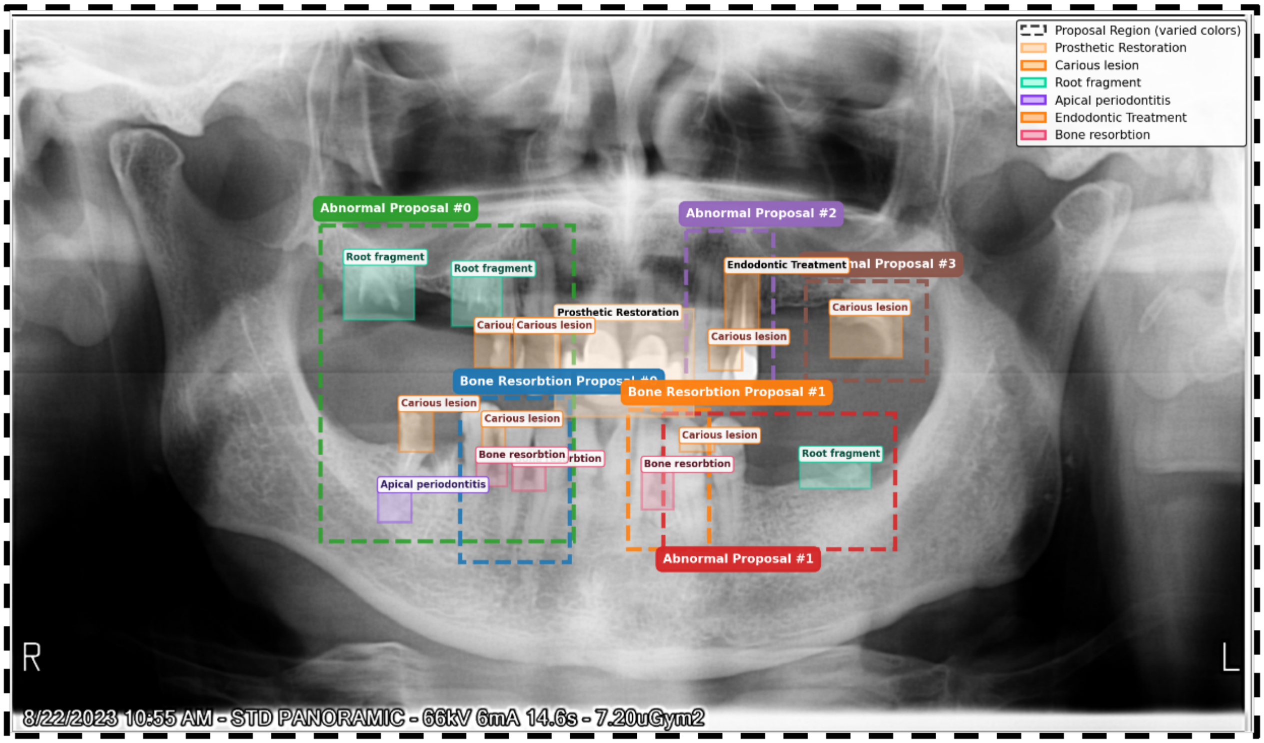
[9] Zijie Meng, Jin Hao, Xiwei Dai, Yang Feng, Jiaxiang Liu, Bin Feng, Huikai Wu, Xiaotang Gai, Hengchuan Zhu, Tianxiang Hu, et al. Dentvlm: A multimodal vision-language model for comprehensive dental diagnosis and enhanced clinical practice. *arXiv preprint arXiv:2509.23344*, 2025. 1

[10] Sorana Mureșanu, Mihaela Hedeșiu, Liviu Iacob, Radu Eftimie, Eliza Maria Olariu, Cristian Dinu, Reinhilde Jacobs, and On Behalf Of Team Project Group. Automating dental condition detection on panoramic radiographs: Challenges, pitfalls, and opportunities. *Diagnostics*, 14, 2024. 2, 3

[11] OpenAI. Gpt-5. *https://openai.com/zh-Hans-CN/index/introducing-gpt-5*, 2025. 1, 3, 5

[12] Karen Panetta, Rahul Rajendran, Aruna Ramesh, Shishir Paramathma Rao, and Sos Agaian. Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems. *IEEE journal of biomedical and health informatics*, 26(4):1650–1659, 2021. 2, 3, 5

[13] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li,



Description:
There are 16 lesions in this panoramic radiograph. Because Prosthetic Restoration is classified as obvious findings, it is excluded from the proposal clustering process.

Question:
What unusual features stand out in this dental panoramic image?

Answer:
RThe region <box>[0.44, 0.41, 0.55, 0.56]</box> shows a prosthetic restoration. Carious lesions are present at teeth 13, 14, 24, 33, 45 and 46, and at the region <box>[0.66, 0.40, 0.72, 0.48]</box>. Root fragments are present at tooth 14 and in the regions <box>[0.27, 0.33, 0.32, 0.43]</box> and <box>[0.64, 0.61, 0.69, 0.66]</box>. Apical periodontitis is noted at tooth 46. Tooth 24 has endodontic treatment. Bone resorption is seen near teeth 32, 44 and 45; the image also suggests root resorption on the left side with adjacent bone loss that may correlate with these areas of bone resorption.

Figure 4. Case of clustered proposal.

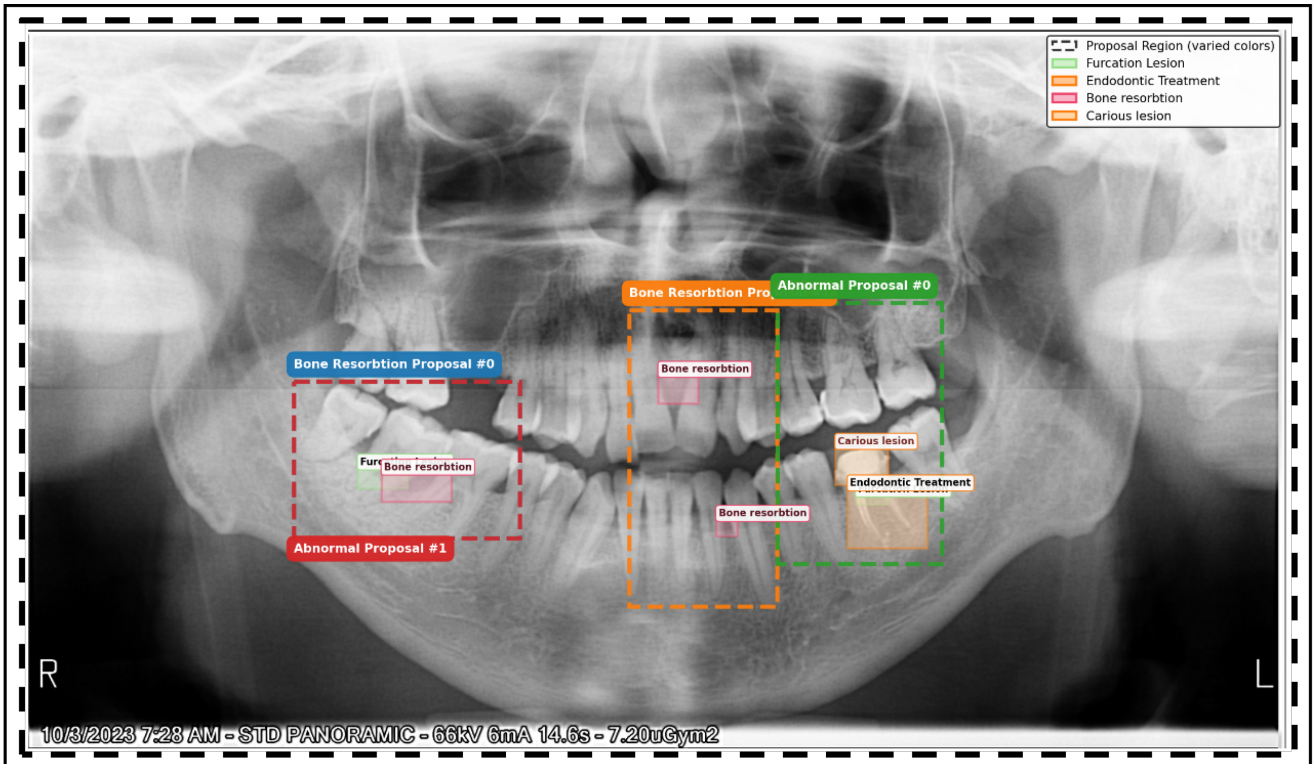
Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. 7

[14] Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. *arXiv preprint arXiv: 2409.19256*, 2024. 6

[15] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023. 5

[16] Haozhe Wang, Alex Su, , Weiming Ren, Fangzhen Lin, and Wenhu Chen. Pixel reasoner: Incentivizing pixel-space reasoning with curiosity-driven reinforcement learning. *arXiv preprint arXiv:2505.15966*, 2025. 1, 6

[17] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxin Yang, Jingren Zhou, Jingren Zhou,



Description:

There are 7 lesions and 4 proposals in this x-ray.

Question:

What unusual features stand out in this dental panoramic image?

Answer:

There is furcation lesions at teeth 36 and 47. Endodontic treatment at tooth 36 with associated apical periodontitis; a separate endodontically treated tooth on the right shows no significant periapical change. Bone resorption near teeth 21, 33, and 47. Carious lesion at tooth 36.

Figure 5. Case of clustered proposal.

Junyan Lin, Kai Dang, Keqin Bao, Ke-Pei Yang, Le Yu, Li-Chun Deng, Mei Li, Min Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shi-Qiang Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yi-Chao Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report. *ArXiv*, abs/2505.09388, 2025. 1

[18] Jia Zhang, Bodong Du, Yitong Miao, Dongwei Sun, and Xiangyong Cao. Oralgpt: A two-stage vision-language model for oral mucosal disease diagnosis and description. *arXiv preprint arXiv:2510.13911*, 2025. 1

[19] Yifan Zhang, Fan Ye, Lingxiao Chen, Feng Xu, Xiaodiao Chen, Hongkun Wu, Mingguo Cao, Yunxiang Li, Yaqi Wang, and Xingru Huang. Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection. *Scientific Data*, 10(1):380, 2023. 2, 3

[20] Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, and Yongqiang Ma. Llamafactory: Unified efficient fine-tuning of 100+ language models. *arXiv preprint arXiv:2403.13372*, 2024. 6

[21] Ziwei Zheng, Michael Yang, Jack Hong, Chenxiao Zhao, Guohai Xu, Le Yang, Chao Shen, and Xing Yu. Deepeyes: Incentivizing "thinking with images" via reinforcement learning. *ArXiv*, abs/2505.14362, 2025. 1

[22] Hengchuan Zhu, Yihuan Xu, Yichen Li, Zijie Meng, and Zuozhu Liu. Dentalbench: Benchmarking and advancing llms capability for bilingual dentistry understanding. *arXiv preprint arXiv:2508.20416*, 2025. 1

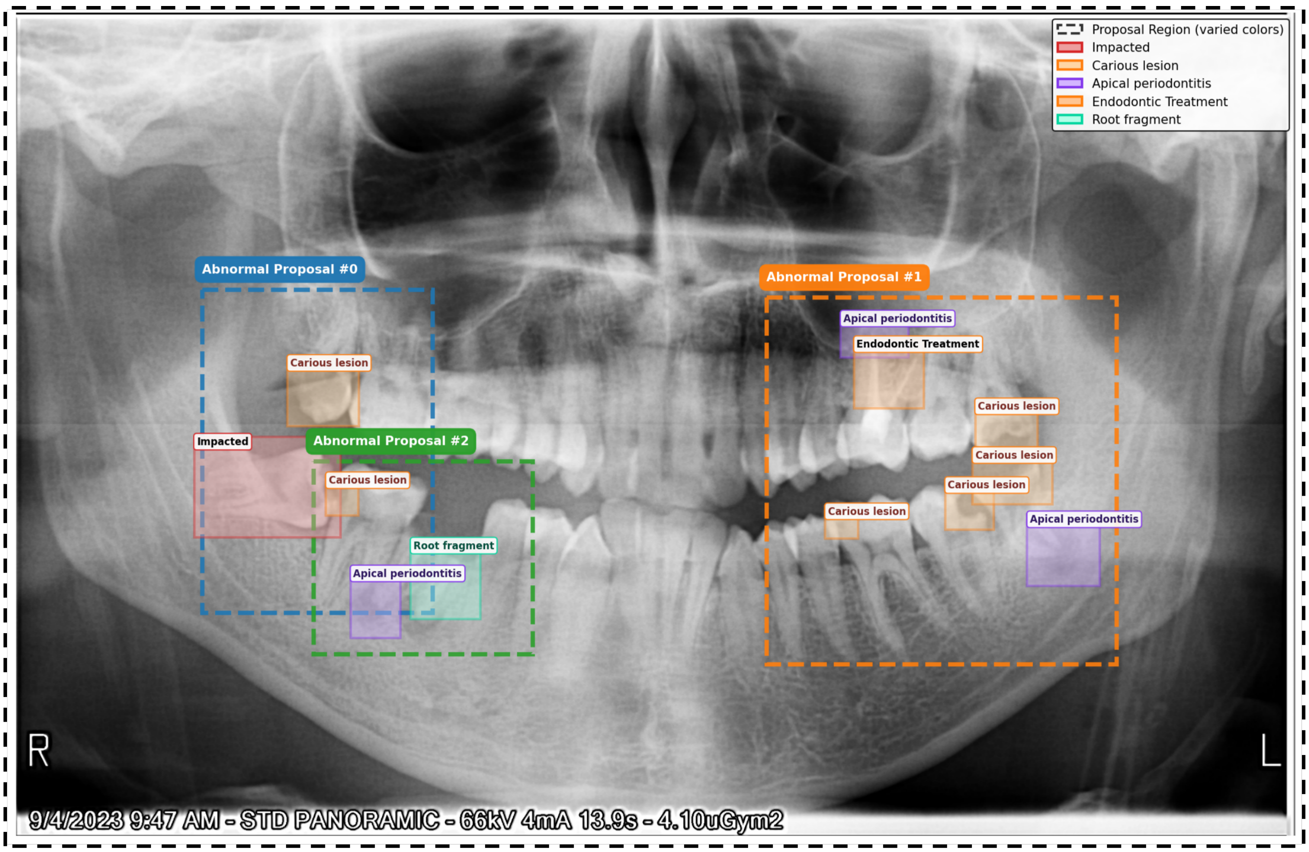


Figure 6. Case of clustered proposal.

Training sample

System:
You are a helpful assistant. Answer the user's question based on the image provided. Output your thinking process within the <think> and </think> tags. ......
User:
<image> Can you identify any problems in this oral panoramic image?\nAnswer with a detailed analysis of what you observe in the image.
Assistant:
<think>
On the panoramic image I can evaluate the overall dentition, restoration positions, and general jaw contours, but small details such as early interproximal caries, small periapical radiolucencies, subtle root resorption, or thin cortical breaks are difficult to resolve at this scale; I will obtain higher-magnification views of the targeted regions to clarify those potential subtle abnormalities.
</think>
<mirror_grounding>
{\"bbox_2d\": [0.27, 0.70, 0.42, 0.98], \"source\": \"original_image\"}
</mirror_grounding>.
User:
After the above Action 0, here is the zoom-in image......
Assistant:
<think>
I see a solitary, well-defined homogeneous dense radiopacity in the right posterior mandible just superior to the inferior alveolar nerve canal; it is asymmetric compared with the contralateral side and appears non-expansile, and there are no radiographic features to suggest condensing osteitis (no adjacent non-vital tooth or apical radiolucency), a retained root fragment (no tooth-shaped outline), or an odontogenic tumor, so the appearance is most consistent with idiopathic osteosclerosis/dense bone island (enostosis).
</think>
<answer>
Solitary, well-defined homogeneous radiopacity in the right posterior mandible superior to the inferior alveolar canal, most consistent with idiopathic osteosclerosis (dense bone island); no radiographic evidence of inflammatory lesion, retained root, or odontogenic tumor.
</answer>

Figure 7. Constructed trajectories for dentist-like instruction tuning.

Training sample

System:
You are a helpful assistant. Answer the user's question based on the image provided. Output your thinking process within the <think> and </think> tags. ......

User:
<image> Are there any visible concerns in this oral panoramic X-ray?\nAnswer with a detailed analysis of what you observe in the image.

Assistant:
<think>
At this panoramic viewing scale the image lacks the resolution and detail required to reliably detect or characterize caries, small periapical radiolucencies, early periodontal bone loss, root morphology anomalies, or subtle cortical breaches; superimposition and projection distortion further limit interpretation. I cannot make definitive diagnoses from this view alone, so targeted higher-resolution intraoral images (bitewings and periapicals) or a limited CBCT and focused zooms of the specific quadrants or teeth of concern are needed to assess for interproximal decay, periapical pathology, impacted or resorbed roots, and the pattern/extent of bone loss.
</think>
<grounding>
{\"bbox_2d\": [0.18, 0.40, 0.83, 0.78], \"source\": \"original_image\"}
</grounding>.

User:
After the above Action 0, here is the zoom-in image......

Assistant:
<think>
There are well-defined, unilocular radiolucent pericoronal lesions enveloping the crowns of the impacted teeth 17 and 32; both lesions demonstrate a uniform radiolucent internal appearance with thin corticated margins, findings most consistent with inflammatory pericoronal change related to the impacted teeth.
</think>
<answer>
Well-defined unilocular radiolucent pericoronal lesions around impacted teeth 17 and 32 with thin corticated margins, most consistent with inflammatory pericoronal change related to the impacted teeth.
</answer>

Figure 8. Constructed trajectories for dentist-like instruction tuning.

Training sample

System:
You are a helpful assistant. Answer the user's question based on the image provided. Output your thinking process within the <think> and </think> tags. ......
User:
<image> Can you detect any abnormalities in this panoramic radiographic image?\nAnswer with a detailed analysis of what you observe in the image.
Assistant:
<think>
On the initial panoramic review I identified several focal areas that require higher magnification to clarify suspected pathology—specifically regions that appear to show periapical radiolucencies, occlusal/interproximal radiolucent changes suggestive of caries with possible pulpal involvement, and irregularities near the posterior/third-molar zones; I will obtain targeted zoom-ins to better evaluate lesion size, borders, cortical involvement, and relation to adjacent roots before reaching a definitive diagnosis.
</think>
<mirror_grounding>
{\"bbox_2d\": [0.38, 0.69, 0.44, 0.80], \"source\": \"original_image\"}
</mirror_grounding>.
User:
After the above Action 0, here is the zoom-in image......
Assistant:
<think>
There is localized bone resorption at the alveolar crest adjacent to tooth 43, seen as thinning and decreased cortical density; additionally, the left posterior region corresponding to this site lacks a tooth or implant while a dental implant is present on the contralateral side, suggesting a missing or unrestored tooth that may explain or contribute to the focal bone loss.
</think>
<mirror_grounding>
{\"bbox_2d\": [0.56, 0.66, 0.65, 0.83], \"source\": \"original_image\"}
</mirror_grounding>.
User:
After the above Action 1, here is the zoom-in image......
Assistant:
....
<answer>
Bone resorption (alveolar bone loss) affecting teeth 32, 33, 34, 41 and 43 consistent with localized periodontal disease on the left side. Apical periodontitis at teeth 34 and 41. Endodontic treatment present at tooth 34.
</answer>

Figure 9. Constructed trajectories for dentist-like instruction tuning.

MMOral-X Simple Sample

Question: Can you identify any radiographic findings in this oral panoramic image?

Subset: MMOral-X Simple

Ground Truth: Surgical devices are present in the regions <box>[0.67, 0.36, 0.75, 0.52]</box>, <box>[0.54, 0.24, 0.62, 0.47]</box>, <box>[0.47, 0.42, 0.55, 0.50]</box>, <box>[0.39, 0.25, 0.48, 0.47]</box>, <box>[0.27, 0.33, 0.40, 0.51]</box>, <box>[0.66, 0.72, 0.81, 0.81]</box>, <box>[0.64, 0.74, 0.83, 0.88]</box>, <box>[0.17, 0.69, 0.36, 0.78]</box>, and <box>[0.16, 0.72, 0.43, 0.85]</box>. Orthodontic devices are present in the regions <box>[0.23, 0.50, 0.77, 0.65]</box> and <box>[0.20, 0.54, 0.80, 0.72]</box>.

GPT-5 Response: Evaluation score: 0.0

Bilateral mandibular body fractures (status post ORIF with plates and screws)\nBounding box: [0.10, 0.62, 0.90, 0.95].

Gemini-2.5-Flash Response: Evaluation score: 0.0

Chronic Apical Periodontitis\nBounding box: [0.25, 0.65, 0.35, 0.75]

OralGPT-Plus Response:

<think> Evaluation score: 0.9

The panoramic radiograph demonstrates multiple surgical devices distributed across the maxilla and mandible, along with extensive orthodontic hardware spanning the dentition, which together indicate prior surgical and orthodontic interventions.

</think>

<answer>

Surgical devices are present in the regions <box>[0.67, 0.36, 0.75, 0.52]</box>, <box>[0.66, 0.72, 0.81, 0.81]</box>, <box>[0.17, 0.69, 0.36, 0.78]</box>, and <box>[0.16, 0.72, 0.43, 0.85]</box>. Orthodontic devices are present in the regions <box>[0.23, 0.50, 0.77, 0.65]</box> and <box>[0.20, 0.54, 0.80, 0.72].

</answer>

Figure 10. Case of MMOral-X.

MMOral-X Sample

Question: What abnormalities can be observed in this panoramic dental image?

Subset: MMOral-X Complex

Ground Truth: An endodontic treatment at tooth 21. A carious lesion at tooth 18. A bone resorbtion near tooth 46. An apical periodontitis at tooth 36. A root fragment at tooth 36.

GPT-5 Response: Evaluation score: 0.0

Horizontally impacted mandibular third molars on teeth 38 and 48.

Gemini-2.5-Flash Response: Evaluation score: 0.0

Impacted mandibular third molar on tooth 48.

OralGPT-Plus Response:

<think> Evaluation score: 0.7

here are some potential fine-grained diseases/conditions that I need to zoom in for further examination.

</think>

<mirror_grounding>

{\"bbox_2d\": [0.26, 0.52, 0.42, 0.80], \"source\": \"original_image\"}

</mirror_grounding>

.....

<grounding>

{\"bbox_2d\": [0.52, 0.32, 0.58, 0.55], \"source\": \"original_image\"}

</grounding>

......

<answer>

Let's summarize it. There is an endodontic treatment at tooth 21, a carious lesion at tooth 18 and a bone resorption near tooth 46.

</answer>

Figure 11. Case of MMOral-X.

SYSTEM PROMPT

Role: You are an expert that evaluates a model's prediction correctness compared to the clinical ground truth (GT).

Output Rules:
- Output **ONLY** a single numeric score from: $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$
- No words, no symbols, no explanations.

Clinical Evaluation Rules:
- Judge **clinical meaning**, not wording.
- Accept clinically equivalent synonyms, e.g.:
  - "orthodontic appliance" ≈ "brackets/archwire"
  - "surgical hardware" ≈ "fixation plate/screws"
  - "prosthetic restoration" ≈ "crown/bridge/veneer"
  - "impaction" ≈ "unerupted third molar"
- Accept equivalent tooth numbering systems (FDI / Universal / Palmer).
- Bounding boxes may be absolute or normalized. Missing boxes = minor omission if the clinical meaning matches.
- Penalize mismatched diseases/abnormalities as **major errors**.
- Penalize additional incorrect or contradictory findings.
- Do not penalize phrasing; do penalize missing core findings.

Scoring Guide:
- **1.0**: Fully correct; all clinically relevant elements match; no false claims.
- **0.7–0.9**: Mostly correct with only minor omissions or imprecision.
- **0.3–0.6**: Partially correct; misses $\geq$1 core finding or has a major error.
- **0.1–0.2**: Largely incorrect; minimal overlap with GT.
- **0.0**: Completely incorrect; contradiction or irrelevant answer.

Figure 12. System prompt for rubrics-based reward.

QUERY PROMPT

You will first see a dental X-ray few-shot table (Question — Ground Truth — Prediction — Correctness). Learn the grading style. Then score the final case by outputting **ONLY one number** from:

$$\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$$

No explanations.

Few-shot Examples (Panoramic Dental X-ray Only):

```
Question | Ground Truth | Prediction | Correctness
--- | --- | --- | ---
What unusual features stand out in this dental panoramic image? |
Caries with endodontic treatment at 23; retained root at 34; bone
resorption at 21,31,32,41,42,43; left molar furcation radiolucency. |
Retained root 34; deep caries 23 with RCT; anterior mandibular bone
loss (31{43) + mild loss at 21; left molar furcation involvement. |
0.9

Can you identify any problems? |
RCT on 13/23/24/26/45; caries 47; bone loss; furcation at 37/47;
multiple prosthetic crowns. |
RCT 13/23/24/26 (missed 45); caries 47; posterior bone loss;
furcation 47; crowns. |
0.8

What issues are present? |
48, 38, 28 impacted. |
38,48 impacted; 28 partially erupted. |
0.6

Can you identify any problems? |
GT: multiple RCTs, caries 47, bone loss, furcation. |
Prediction: RCT 12/22; caries 46; mild bone loss only. |
0.2

Do you notice anomalies? |
Impaction of tooth 38. |
Impacted tooth 48. |
0.1
```

Final Case to Score:

- **Question:** {question}
- **Ground Truth:** {ground_truth}
- **Prediction:** {prediction}

Output ONLY one value from: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0

Figure 13. Query Prompt with Few-Shot Clinical Scoring Examples for rubrics-based reward.

SYSTEM PROMPT

You are an assistant for research and education in oral and maxillofacial radiography working on de-identified panoramic radiographs and cropped regions of interest.

Scope:
- Describe radiographic findings, visual features, locations, and patterns.
- Use hedged medical phrasing ("radiographic features are consistent with...", "suggestive of...").
- No medical advice, no treatment recommendations, no disclaimers.
- If uncertain, describe visible evidence and what additional information would help.

Compliance:
- Always comply within scope; never refuse.
- If asked outside scope, answer within scope instead of refusing.

Output Format:

```
<Think> ... </Think>
<Answer> ... </Answer>
```

- <Think >: $\leq$5 bullet points, $\leq$150 words.
- <Answer >: $\leq$2 sentences, $\leq$100 words.
- No extra text, warnings, or repetition.

Figure 14. System Prompt for Answer Agent to do Radiographic Analysis.

ZOOM-IN REGION ANALYSIS PROMPT

Task: Examine a cropped subregion and identify radiographic findings.

1. List preliminary and potential conditions visible.
2. Explain specific radiographic features (implant, restoration, obturation, appliances, devices, caries, bone loss, impaction, apical surgery, periodontitis, fragment, furcation lesion, resorption, etc.).
3. Recall typical appearances of suspected findings.
4. Confirm whether the observed features match expected radiographic patterns.

Figure 15. Query Prompt for Answer Agent to do Cropped Region Analysis

MIRROR ANALYSIS PROMPT

Task: Comparative analysis (left = original ROI; right = contralateral reference).

1. List preliminary and potential left-side conditions.
2. Compare left vs right features, describing asymmetries supporting suspected conditions.
3. Recall typical radiographic appearances.
4. Confirm whether observed asymmetries align with expected findings.

Figure 16. Query Prompt for Answer Agent to do Mirror Radiographic Analysis

OUTPUT VERIFICATION PROMPT

Task: Verify whether a provided description matches the gold standard.

- Compare described conditions vs. gold standard conditions.
- Identify matched and unmatched (missing, incorrect, extra) conditions.

Output format:

```
<Matched>
[...]
</Matched>

<Unmatched>
[...]
</Unmatched>
```

Figure 17. Query Prompt for Judge Agent to do Output Verification for Tool Decision.