EAGLE-360: Embodied Active Global-to-Local Exploration in 360°

Anonymous authors
Paper under double-blind review

Abstract

Precise target localization in 360° environments requires integrating global spatial context with fine-grained visual evidence while accounting for projection distortion and horizontal wrap-around geometry. For multimodal large language models (MLLMs), the challenge is not only to select informative views but also to derive geometrically consistent estimates of the target direction from visual observations. We present **EAGLE-360**, an Embodied Active Global-to-Local Exploration framework that leverages panoramic context to guide iterative, tool-augmented target localization. We introduce Spherical Atlas Rotary Position Embedding (AtlasRoPE) for local spherical positional modeling, alongside Spherically Stratified Replay (SSR) and Spherical Direction Regularization (SDR) for geometry-aware training. This training strategy increases exposure to underrepresented directions and incorporates continuous spherical supervision, explicitly accounting for seam-adjacent and high-latitude targets without adding inference-time modules. To support this framework, we construct a curated panoramic search dataset with multi-turn trajectories interleaving reasoning, projection actions, and visual observations. We train the model through supervised fine-tuning (SFT) followed by reinforcement learning based on Group Relative Policy Optimization (GRPO) to develop its multi-turn search capabilities. Experiments show that EAGLE-360-8b leads the main comparison on EAGLE-360 with 84.01% localization accuracy, together with strong transfer to H*Bench, achieving 75.00% object-search success under panoramic initialization without benchmark-specific training.

1 Introduction

Target localization in 360° panoramas connects scene understanding with precise viewing directions. Language-guided search and navigation link linguistic goals to visual environments (Qi et al., 2020; Khanna et al., 2024). Multimodal large language models (MLLMs) offer a route to generalist navigation (Zheng et al., 2024). Panoramas reveal global layout; perspective views expose details difficult to resolve at a fixed panoramic resolution. Effective localization requires deciding where to inspect and when to answer.

Hstar introduces HVS-3B for active search through local observations (Yu et al., 2026); PanoWorld learns from full equirectangular projections (ERPs) (Wang et al., 2026). Combining global and local evidence requires handling ERP distortion at the poles and discontinuity at the horizontal seam. Training data also concentrate near the horizon, while coordinate-token supervision does not encode angular proximity. These challenges motivate geometry-aware positional encoding, directional coverage and supervision.

We present **EAGLE-360**, an Embodied Active Global-to-Local Exploration framework for panoramic target localization (Figure 1). Given a panorama and a query, the model generates chain-of-thought (CoT) text followed by either a final direction or a projection request specifying azimuth, elevation and field of view (FoV). Returned views support refinement and confirmation. The policy can answer directly or acquire up to three views within the same panorama.

Three components support this interface. Building on rotary position embedding (RoPE) and its horizontal rolling variant (Ren et al., 2025), **Spherical Atlas Rotary Position Embedding (Atlas-**

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



Figure 1: **EAGLE-360: global context and optional local inspection.** Equirectangular projection (ERP) distortion, seam continuity and uneven directional coverage motivate AtlasRoPE, SSR and SDR. The model either submits a direction or requests a perspective view to inspect the target before answering. The six SSR strata cover central-left/right, seam-left/right and north/south polar directions; only the replay pool has equal quotas. Arrows and coverage shading illustrate the design, not a scored rollout.

RoPE) assigns local spherical charts to tokens of the initial ERP image, adjusting rotary phases in the language decoder around seam-adjacent and polar regions.

Spherically Stratified Replay (SSR) increases exposure to underrepresented directions through geometrically consistent image and annotation transformations. **Spherical Direction Regularization (SDR)** supervises a direction predicted from the decoder state preceding the current response, using an auxiliary head discarded at inference. Together, they address positional modeling, training coverage and continuous directional supervision.

We construct 21,884 geometry-guided reference trajectories from 4,882 source panoramas, interleaving reasoning, projection actions and observations. Stage-specific supervised fine-tuning (SFT) establishes single-turn localization and CoT before teaching tool decisions and observation-conditioned replies. We adapt Group Relative Policy Optimization (GRPO) for reinforcement learning (RL), using localization rewards and a bounded tool-use bonus to train initial decisions and subsequent responses. Our 713-question direction-balanced test uses panorama-disjoint source splits and consistent image–annotation reorientation to evaluate horizontal and polar localization.

EAGLE-360-8b leads the main comparison with **84.01%** accuracy and **12.03**° mean angular error (Table 1). On H*Bench, it achieves **75.00%** object-search and **58.90%** overall success, leading our comparison under panoramic initialization without benchmark-specific training (Table 2). Matched SFT ablations show substantial polar gains from SSR and additional gains from AtlasRoPE and SDR with replay (Table 3). Feedback controls demonstrate the value of real observations for confirmation and completion, especially at limited initial resolution (Table 4).

Contributions. (1) A global-to-local localization framework trained through progressive SFT and GRPO-derived reinforcement learning. (2) AtlasRoPE, SSR and SDR for spherical positional modeling, direction-stratified replay and continuous supervision. (3) A panoramic search corpus and direction-balanced evaluation, with matched component ablations, controlled feedback experiments and H*Bench transfer results.

2 Related Work

Panoramic understanding and search. OSR-Bench, ODI-Bench and OmniCoT evaluate omnidirectional reasoning (Dongfang et al., 2025; Yang et al., 2025; He et al., 2026). Dense360 supplies grounding supervision; PanoWorld combines panorama-native tuning with spherical cross-

attention (Zhou et al., 2025; Wang et al., 2026). 360-R1 and PanoEnv learn panoramic reasoning through RL (Zhang et al., 2025; Lin & Zheng, 2026). Active-view precedents include Refer360, Hstar and Imagining360 (Cirik et al., 2020; Yu et al., 2026; Zhang et al., 2026). EAGLE-360 combines full-ERP context with optional perspective inspection and a global directional answer. Actions change orientation within a fixed panorama, unlike navigation between locations (Qi et al., 2020; Khanna et al., 2024).

Geometric representations and positional encoding. Spherical convolutional networks, PanoSwin and SphereUFormer incorporate spherical geometry into correlations or attention (Cohen et al., 2018; Ling et al., 2023; Benny & Wolf, 2025). RoPE encodes positions through query/key rotations (Su et al., 2021). PanoSplatt3R's RoPE Rolling models horizontal periodicity; Dense360's ERP-RoPE addresses continuity and latitude-dependent spacing; FishRoPE adapts rotary coordinates to fisheye geometry (Ren et al., 2025; Zhou et al., 2025; Ahuja et al., 2026). AtlasRoPE uses overlapping local spherical charts with smooth transitions to native coordinates for tokens of the initial ERP image in the language decoder. It changes positional phases without resampling features, complementing geometric features introduced in the visual stream, as in PanoWorld.

Directional coverage and supervision. Pano Stretch augments panoramic geometry, class-balanced learning reweights underrepresented categories, and geometric multi-task objectives complement semantic supervision (Sun et al., 2019; Cui et al., 2019; Kendall et al., 2018). SSR instead adds direction-stratified replay with consistent image–annotation transformations. SDR supplies continuous spherical supervision through a training-only direction head. Equal-count replay controls separate directional coverage from training volume.

Visual tools and policy learning. ReAct interleaves reasoning, actions and observations; V* and CogCoM extend visual reasoning through search and manipulation (Yao et al., 2023; Wu & Xie, 2024; Qi et al., 2025). Pixel Reasoner, DeepEyes, VTool-R1 and ACTIVE-o3 learn visual tool use through reinforcement learning (Wang et al., 2025a; Zheng et al., 2026; Wu et al., 2026; Zhu et al., 2026). We specialize this interface to spherical projections using progressive SFT and an objective derived from GRPO (Shao et al., 2024); our adaptation assigns separate advantages to initial responses and subsequent continuations. Following the distinction between intrinsic and tool-induced performance (Ma et al., 2026), fixed-policy feedback controls assess the utility of observations.

3 Method

EAGLE-360 combines an interactive localization policy with geometry-aware representation and training (Figure 2). The same model chooses whether to answer or inspect a view; AtlasRoPE, SSR and SDR supply complementary geometric structure.

3.1 Task and Interaction

Given an ERP image and a query, the policy generates CoT text followed by exactly one projection-tool call or a final direction $(\widehat{\theta}, \widehat{\phi})$. Azimuth $\theta \in [-\pi, \pi)$ and elevation $\phi \in [-\pi/2, \pi/2]$ define the unit ray $\mathbf{d}(\theta, \phi) = (\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi)$. A call specifies camera azimuth, elevation and FoV. Equirectangular-to-perspective rendering returns a view processed by the shared vision encoder; the view and camera metadata join the dialogue. Inference permits at most three calls and four assistant responses, ending with an explicit final answer. No ground-truth information enters tool feedback.

We score the submitted direction using the annotated bounding field of view (BFoV). For predicted ray $\widehat{\mathbf{d}}$, ground-truth ray $\mathbf{d}^*$ and annotated angular diagonal δ_{bbox}, the hit criterion is

$$d_{\mathrm{gc}} = \arccos\Big(\mathrm{clip}_{[-1,1]}(\widehat{\mathbf{d}}^\top \mathbf{d}^*)\Big) \le r, \qquad r = \delta_{\mathrm{bbox}}/2. \tag{1}$$

Here d_{gc} is the great-circle distance (GCD), r is the angular acceptance radius, and $\mathrm{clip}_{[a,b]}$ clamps a value to $[a, b]$. BFoV accuracy uses this circular tolerance rather than rectangular containment. Geometric equations use radians unless stated otherwise; tool arguments and reported errors use degrees. Invalid final answers count as incorrect with 180° error, without substituting the camera center. Appendices B and D.1 provide implementation details.

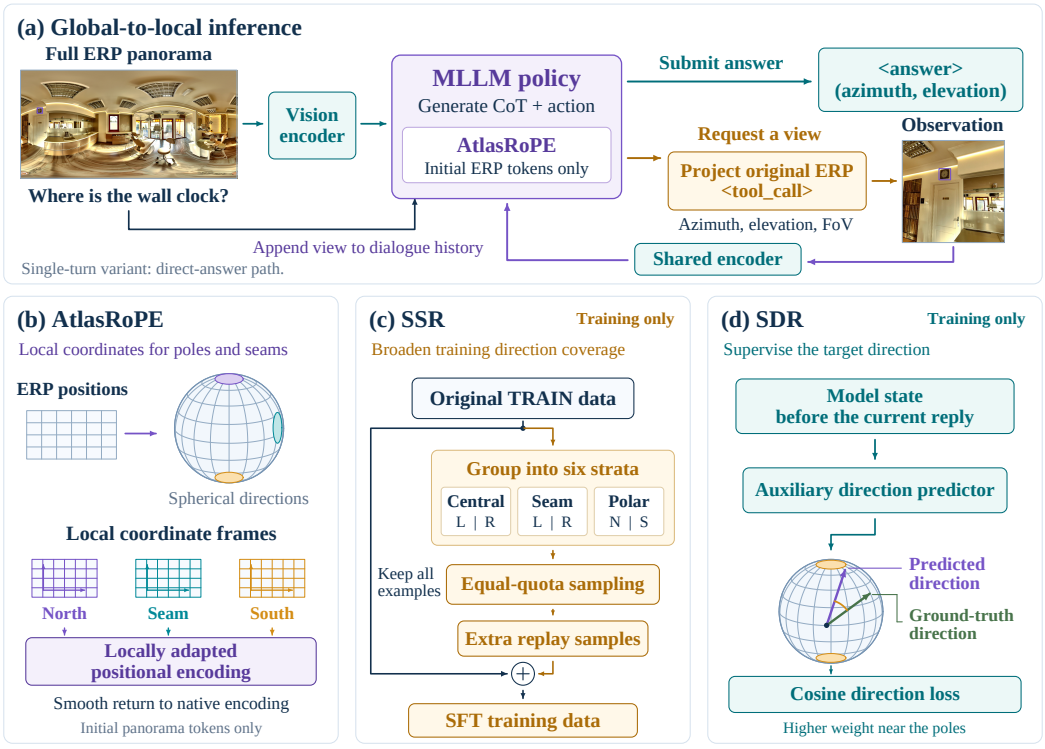


Figure 2: **EAGLE-360 framework and geometry-aware training. (a)** The policy answers directly or requests a perspective image, which joins the dialogue history. **(b)** AtlasRoPE adjusts rotary phases in the language decoder for tokens of the initial ERP image through local spherical charts. **(c)** SSR adds equal-quota replay from six spherical strata to ordinary training data. **(d)** SDR predicts the target direction from the decoder state immediately before the supervised reply, including its CoT; a latitude-weighted cosine loss trains this auxiliary head, which is removed at inference.

3.2 Spherical Atlas Rotary Position Embedding (AtlasRoPE)

Horizontal RoPE Rolling (Ren et al., 2025) changes the seam location across heads. AtlasRoPE introduces local spherical frames around the seam and poles, modifying the language decoder's multimodal rotary position embedding (MRoPE) only for tokens of the initial ERP image (Bai et al., 2025). Text, temporal channels, the vision encoder and perspective tool views retain the backbone's original (native) encoding.

Each visual token i at merged-grid coordinates (x_i, y_i) defines a spherical ray. Expressing it in rotated frame k and mapping its local angles back to grid units gives $(\widetilde{x}_{ki}, \widetilde{y}_{ki})$. A smooth gate $g_{ki} \in [0, 1]$ activates this chart within a spherical cap. For native phase $\psi_{i\ell}^{\text{native}}$ at rotary frequency ω_ℓ,

$$\psi_{ki\ell} = \psi_{i\ell}^{\text{native}} + \eta g_{ki} \omega_\ell \Delta_{ki\ell}, \qquad \Delta_{ki\ell} = \begin{cases} \widetilde{y}_{ki} - y_i, & \ell \in \mathcal{I}_h, \\ \widetilde{x}_{ki} - x_i, & \ell \in \mathcal{I}_w, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where ℓ indexes rotary frequencies, $\mathcal{I}_h, \mathcal{I}_w$ index height/width channels, and $\eta \in [0, 1]$ controls the phase adjustment (1 in the matched component study). Each key/value group receives one chart with identical phases for its keys and associated queries. One group stays native; other charts return smoothly to native phases outside their caps. AtlasRoPE changes positional phases without resampling features or adding trainable parameters. Appendix B.1 specifies the frames, anchors and gates.

A polar direction can lie near the equator of a pole-centered chart, reducing local coordinate distortion. Bounded chart support and a retained native group preserve the original positional reference; perspective inspection supplies additional visual evidence.

3.3 Geometry-Aware Training

Spherically Stratified Replay (SSR). We retain 21,884 ordinary training records and add 469 replay records from each of six strata: central-left/right, seam-left/right and north/south polar. The resulting 24,698-record pool increases seam and polar exposure. Images, targets and coordinate-bearing supervision are transformed consistently. Equal quotas apply only to the initial replay pool; the combined data and later scenario sampling need not be uniform. These training strata differ from the evaluation bands. Controls replace SSR with equal-count ordinary replay under matched budgets (Appendix B.2).

Yaw augmentation preserves elevation; spherical reorientation also exposes targets to high-latitude ERP distortion. Rotating image rays, target directions and camera parameters together preserves their geometric relationship. Replay broadens directional exposure without additional annotations, while the matched-count control separates coverage from training volume.

Spherical Direction Regularization (SDR). A training-only linear head predicts a normalized Cartesian direction $\widehat{\mathbf{d}}_i$ from the last decoder state *before the first supervised token* of the current reply. It sees prior observations and dialogue, but not the teacher-forced current CoT or answer. For a batch of B examples, ground-truth ray $\mathbf{d}_i^* = \mathbf{d}(\theta_i^*, \phi_i^*)$ and scenario weight w_i (1 when unweighted),

$$\mathcal{L}_{\mathrm{SDR}} = \frac{1}{B} \sum_{i=1}^{B} w_i (1 + \sin^2 \phi_i^*) \left[1 - \mathrm{clip}_{[-1,1]}(\widehat{\mathbf{d}}_i^\top \mathbf{d}_i^*) \right], \tag{3}$$

$$\mathcal{L}_{\mathrm{SFT}} = \mathcal{L}_{\mathrm{CE}} + 0.02 \mathcal{L}_{\mathrm{SDR}}. \tag{4}$$

The cross-entropy loss $\mathcal{L}_{\mathrm{CE}}$ averages each example's supervised-token loss with the same scenario weights. The cosine penalty encodes angular proximity across the azimuth seam, and its latitude weight grows from 1 at the equator to 2 at the poles. The auxiliary head is discarded at inference; final coordinates remain language-generated. Appendix B.3 specifies normalization and gradient handling.

For unit directions, the penalty equals one minus the cosine of their angular separation, so nearby directions receive similar supervision even across the azimuth seam. AtlasRoPE, SSR and SDR respectively modify positional encoding, training exposure and supervision, while retaining the same inference interface.

3.4 Progressive Policy Optimization

Supervised fine-tuning (SFT). Training progresses from single-turn localization and CoT to tool selection, observation-conditioned responses and termination. Only the current assistant response—CoT, tool call or final answer, and end token—receives language supervision; prompt and history tokens are masked. Later-turn records include preceding actions and returned images without repeatedly supervising the first decision. Direct answers and first-turn replay remain in the mixture to preserve both stopping and continuation behavior. Matched component configurations retain their AtlasRoPE/SSR/SDR settings and stage budgets; the main model receives additional interaction-focused SFT before reinforcement learning. Appendix C specifies the training sequences and budgets.

Group Relative Policy Optimization (GRPO). We adapt the critic-free objective of Shao et al. (2024). Each question yields four sampled first responses (*roots*). A final answer or invalid response terminates; an executed first tool call produces four continuations sharing the returned view. Each root advantage is its mean terminal return centered across roots without standard-deviation normalization. Continuations receive separate advantages for final accuracy and error reduction from the first camera center; invalid endings are penalized.

We replay each root once and average its continuations to avoid overweighting shared decisions. A clipped token-level objective combines Kullback–Leibler (KL) regularization to a frozen SFT reference with auxiliary SDR. A bounded tool-use bonus, gated by successful localization, decreases with the fraction of roots using a tool and does not grow with later calls. Direct answers remain permitted. The camera center is a tool-call parameter, not a submitted answer. Appendix C.3 specifies this GRPO-derived objective and checkpoint selection; the matched component study ends before reinforcement learning.

4 DATASET

Sources and splits. EAGLE-360 combines Matterport3D (Chang et al., 2017), 2D-3D-S (Armeni et al., 2017) and Kuula (Kuula, 2026) panoramas with a curated PanoInfinigen (Bozic et al., 2026) expansion. After removing source-panorama overlap with the original test split, training contains 21,884 reference trajectories covering 5,471 targets on 4,882 panoramas. Four yaw orientations per target broaden horizontal coverage; SSR provides additional polar exposure. Evaluation uses 713 test questions and 319 separate validation questions. We construct a fixed direction-balanced evaluation by jointly reorienting selected images and annotations, retaining one scored orientation per question; rotations do not create independent scenes. Direction assignment uses annotations rather than model predictions. Five direction groups contain 119 questions each and the sixth contains 118, giving all six directions nearly equal weight in the overall score. All compared models receive the same frozen questions and target directions.

Trajectory synthesis and supervision. GroundingDINO (Liu et al., 2025) detections, spherical BFoV checks and semantic screening identify targets. GPT-4o-mini (OpenAI, 2024b) assists description generation and target-uniqueness checks for real panoramas; Qwen3-VL (Bai et al., 2025) screening and human review curate the expansion. Annotated geometry guides coarse-to-fine camera sequences, rendered observations and templated reasoning, producing 88,189 tool replies and 21,884 final replies. Stage-specific SFT instead supervises individual current replies with frozen prefixes: 24,698 records per single-turn stage, 4,589 multi-turn records and 1,316 interaction-focused records. Later CoT sets also include dialogue prefixes generated offline by teacher and student models. These counts describe optimization records, not unique scenes; construction, filtering and provenance are detailed in Appendix A.

Supervision covers initial decisions, refinement, search recovery, confirmation and termination. Student-generated prefixes expose the policy to imperfect earlier decisions; confirmation examples teach it to answer when the available view suffices. Shared frozen prefixes keep observation histories identical across component configurations.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

We evaluate the direction-balanced test set (713 questions) and select GRPO checkpoints on the direction-balanced validation set (319 questions). H*Bench reports success rates for its object-search (HOS) and path-search (HPS) tasks using the official scorer, with Overall $= 0.6\,\mathrm{HOS} + 0.4\,\mathrm{HPS}$.

Qwen3-VL-8B (SFT), termed Plain SFT, is the multi-turn CoT baseline without AtlasRoPE, SSR or SDR. EAGLE-360-8b additionally uses geometry-aware training, interaction-focused SFT and GRPO. Both start from a panorama and decode greedily, allowing at most three tool calls and 512 tokens per response. Direct-answer baselines use panoramas; HVS-3B retains its native local-view search. Main-table comparisons measure complete systems; matched component and feedback effects are tested separately. Appendix D records model-specific preprocessing, budgets and scoring. Anonymous evaluation code, model weights and test data are listed in Appendix D.3.

Acc. denotes BFoV localization accuracy: a prediction is correct when its great-circle angular error does not exceed half the annotated BFoV angular diagonal (Eq. 1). GCD (great-circle distance) is the mean angular error in degrees. Both include all questions; invalid outputs count as incorrect with 180° error. Direction-wise accuracy applies the same criterion within ground-truth groups: Up/Down ($|\phi| \geq 45^\circ$) and four azimuth bands otherwise. Polar accuracy pools Up/Down by question count. Component comparisons use identical question IDs.

5.2 PANORAMIC TARGET LOCALIZATION

EAGLE-360-8b achieves **84.01% accuracy** and **12.03° GCD**, leading Table 1 and exceeding Plain SFT, GPT-5.6 Sol and PanoWorld by 7.71, 12.48 and 15.01 percentage points. Up/Down accuracy reaches 63.87%/62.71%, versus 38.66%/35.59% for Plain SFT and 16.81%/14.41% for PanoWorld.

Table 1: Balanced EAGLE-360 test (713 questions). Acc.: BFoV accuracy; GCD: mean angular error. Top-three distinct values use red, green, and blue; best values are bold. ‡: tool-enabled; †: provider-internal reasoning.

Method	Acc.	GCD	Direction-wise Acc. (%)↑					
	(%)↑	(°)↓	Front	Back	Left	Right	Up	Down
Samples	713	713	119	119	119	119	119	118
Open-source Models								
InternVL3.5-4B	2.52	102.34	0.00	5.88	2.52	4.20	0.84	1.69
InternVL3.5-8B	5.33	78.46	11.76	4.20	5.04	6.72	1.68	2.54
Qwen3-VL-4B-Instruct	7.57	72.73	20.17	11.76	8.40	5.04	0.00	0.00
Qwen3-VL-8B-Instruct	5.61	78.34	2.52	17.65	6.72	5.88	0.00	0.85
Qwen3.5-4B	9.26	53.96	10.08	13.45	11.76	10.08	7.56	2.54
Qwen3.5-9B	16.41	40.30	27.73	19.33	14.29	17.65	10.08	9.32
Proprietary Models								
GPT-4o	9.12	52.15	9.24	0.00	14.29	25.21	3.36	2.54
Gemini 2.5 Pro†	17.95	34.97	10.08	15.13	26.05	21.85	21.85	12.71
Claude Sonnet 5	17.67	39.90	14.29	47.06	6.72	17.65	11.76	8.47
GPT-5.5	31.84	25.43	34.45	48.74	25.21	51.26	13.45	17.80
Gemini 3.1 Pro Preview†	48.81	15.88	67.23	57.14	42.02	35.29	43.70	47.46
GPT-5.6 Sol	71.53	15.06	89.92	81.51	89.92	89.08	43.70	34.75
Panoramic Comparison Methods								
PanoWorld-9B (public)	69.00	17.68	**94.96**	**96.64**	92.44	98.32	16.81	14.41
HVS-3B (public)‡	14.73	73.05	13.45	27.73	26.89	20.17	0.00	0.00
Our Method								
Qwen3-VL-8B (SFT)‡	76.30	21.42	94.12	93.28	**96.64**	**99.16**	38.66	35.59
EAGLE-360-8b‡	**84.01**	**12.03**	94.12	89.92	95.80	97.48	**63.87**	**62.71**

Horizontal accuracy remains 89.92–97.48%, demonstrating stronger polar coverage alongside high horizontal performance. The direction-wise results explain this overall gain. PanoWorld and Plain SFT already obtain high horizontal accuracy, but their polar scores are substantially lower. EAGLE-360 narrows that disparity, bringing the largest improvements to directions that are scarce in ordinary training. Reporting all six directions thus reveals progress that a horizon-dominated test could obscure.

5.3 Transfer to H*Bench

Our models transfer without H*Bench-specific training, retaining panoramic initialization and projection tools; literature protocols are identified in Table 2. EAGLE-360-8b achieves the highest reported **HOS (75.00%)** and **Overall (58.90%)**, improving on Plain SFT by 8.17 and 5.60 percentage points, and zero-shot PanoWorld by 13.20 and 2.80 points. HPS also increases from 33.00% to 34.75% over Plain SFT. These results demonstrate effective transfer to panoramic object search.

The strongest transfer gains occur in object search, which shares our target-localization objective.

5.4 Geometry-Aware Component Ablations

Table 3 compares six matched multi-turn CoT models before GRPO. At each stage, all configurations use one epoch with matched record counts and update budgets; ordinary replay replaces SSR in controls. EAGLE-360-8b in Table 1 follows a different SFT training sequence (Appendix C).

Table 2: H*Bench success rates (%); best values are bold. †: Overall computed from rounded HOS/HPS. Hstar (Yu et al., 2026) uses local initialization; PanoWorld (Wang et al., 2026) and ours use panoramas. HVS-3B uses benchmark-specific training; PanoWorld and ours transfer without it. Protocols: Appendix D.2.

Method	HOS ↑	HPS ↑	Overall ↑
Open-source Models			
InternVL3.5-4B	3.08	4.81	3.77†
InternVL3.5-8B	6.38	7.25	6.73†
Qwen2.5-VL-3B	14.83	6.44	11.47†
Qwen2.5-VL-7B	11.38	6.31	9.35†
Gemma-3-4B	17.13	14.44	16.05†
Gemma-3-12B	10.21	14.50	11.93†
Kimi-VL-A3B	4.92	4.32	4.68†
Proprietary Models			
GPT-4o	19.75	23.69	21.33†
Gemini 2.5 Pro	31.96	33.00	32.38†
Panoramic Comparison Methods			
HVS-3B (SFT)	40.83	23.00	33.70†
HVS-3B (SFT+RL)	47.38	24.94	38.40†
PanoWorld-9B (zero-shot)	61.8	**47.5**	56.1
Our Method			
Qwen3-VL-8B (SFT)	66.83	33.00	53.30
EAGLE-360-8b	**75.00**	34.75	**58.90**

Table 3: Matched multi-turn CoT component ablations after SFT, before GRPO. Accuracies are percentages; GCD is in degrees. Polar pools Up/Down ($|\phi| \geq 45°$). Best overall accuracy, GCD and polar accuracy are bold.

Configuration	Overall ↑	GCD ↓	Front	Back	Left	Right	Up	Down	Polar ↑
Samples	713	713	119	119	119	119	119	118	237
Plain	76.30	21.42	94.12	93.28	96.64	99.16	38.66	35.59	37.13
AtlasRoPE	75.88	18.62	96.64	91.60	95.80	97.48	34.45	38.98	36.71
Plain + SSR	83.03	13.77	95.80	94.12	97.48	98.32	58.82	53.39	56.12
AtlasRoPE + SSR	84.01	**12.63**	96.64	89.92	96.64	99.16	63.87	57.63	60.76
Plain + SSR + SDR	84.15	12.83	96.64	93.28	96.64	98.32	60.50	59.32	59.92
AtlasRoPE + SSR + SDR	**84.43**	13.48	96.64	91.60	95.80	98.32	63.03	61.02	**62.03**

AtlasRoPE lowers GCD from $21.42°$ to $18.62°$; with SSR, it raises overall/polar accuracy from 83.03%/56.12% to 84.01%/60.76%. Its local frames complement replay exposure to polar and seam-adjacent content. **SSR** gives the largest gains: 6.73–8.13 percentage points overall and 18.99–24.05 in polar accuracy, consistent with increased directional exposure. **SDR** adds 0.42–1.12 percentage points overall on top of SSR. The full SFT configuration reaches 84.43% overall and 62.03% polar accuracy.

SSR exposes targets in less common orientations while preserving image–annotation geometry; its matched-count control fixes training volume. AtlasRoPE's additional polar gains with SSR are consistent with broader exposure to its local charts. SDR adds continuous angular supervision. The full configuration has the highest overall and polar accuracy, whereas AtlasRoPE+SSR has the lowest GCD, showing that hit rate and angular error capture complementary aspects of performance.

5.5 Visual Feedback in Multi-Turn Inference

We freeze EAGLE-360-8b and evaluate all 713 questions with standard or 512×256 initial panoramas, using no tools, real tool views or mismatched tool views from another panorama. Real-view and mismatched-view conditions share the first response and three-call budget; the no-tool condition adds an instruction disabling tools. All retain CoT. Table 4 uses NVIDIA V100 GPUs; its standard real-view result differs by one correct prediction from the main evaluation on RTX PRO 6000 Blackwell GPUs (Appendix D.1).

Table 4: Fixed-model feedback ablation on all 713 questions. Valid: explicit final answers; Calls: total executed tool calls. GCD is in degrees and includes 180° for invalid outputs. All conditions share the evaluation environment.

First ERP	Feedback	Hits	Acc. (%)	GCD (°)↓	Valid	Calls
Standard	No tools	559	78.40	27.22	642	0
Standard	Real tool views	**600**	**84.15**	**12.25**	707	121
Standard	Mismatched tool views	560	78.54	24.62	653	242
Low-resolution	No tools	451	63.25	43.00	597	0
Low-resolution	Real tool views	**510**	**71.53**	**20.70**	697	262
Low-resolution	Mismatched tool views	435	61.01	47.54	571	527

Real tool views outperform mismatched tool views by **5.61 percentage points** with standard input and **10.52 percentage points** with low-resolution input, producing more valid answers with fewer calls. Compared with the no-tool condition, accuracy increases from 78.40% to 84.15% and from 63.25% to 71.53%. The larger low-resolution gain demonstrates the value of local visual evidence for confirmation and successful submission.

The mismatched-view control preserves the projection interface, first action and camera metadata while changing observation content, testing whether relevant evidence helps the fixed policy complete the task. In the low-resolution condition, tool views still come from the full-resolution source and can recover details absent from the initial representation. Local inspection thus supports confirmation and termination, including when the final answer retains an earlier estimate.

6 Conclusion

EAGLE-360 combines panoramic context, optional perspective inspection and geometry-aware learning. The main comparisons show strong localization and transfer, while matched controls identify gains from directional replay, complementary geometric components and relevant visual feedback. The evidence supports a division of labor: panoramic input supplies a global estimate, while targeted views support evidence-based confirmation and completion. Balanced directional evaluation makes the contribution of polar coverage visible alongside overall performance. Future work will target more effective coordinate refinement, uncertainty-aware view selection and compact visual histories for dynamic settings.

Limitations

Evaluation uses static panoramas. Reorientation broadens directional coverage but changes gravity alignment and does not create independent scenes. Main-table systems differ in training, initial views and computation. Experiments use single training runs, and test reuse during development makes independent scenes and additional seeds important for assessing robustness. Feedback controls test observation utility, rather than isolating CoT or guaranteeing progressive coordinate refinement. Tool views add sequential inference cost; physical navigation remains outside this evaluation. Evaluation protocols are provided in Appendix D.

AI USE STATEMENT

We used generative AI and LLM-based assistants in the research and preparation of this paper. The following describes their roles, consistent with the AI-assistance disclosures in the submission form.

Writing and drafting. AI assisted with drafting sections of the manuscript, proposing its organization, translating and rephrasing text, and polishing clarity, grammar and presentation. It also helped prepare tables and format references.

Retrieval and discovery. AI assisted with finding related work, searching for relevant methods and benchmark protocols, and summarizing literature to support comparisons and contextualize the contributions.

Research ideation and execution. AI assisted with brainstorming and refining hypotheses, developing the conceptual framework, designing experiments and ablation controls, suggesting training settings, implementing and debugging methods and data-processing code, and organizing training and evaluation runs. It also supported inspection of model trajectories, qualitative error analysis and interpretation of experimental results.

Synthetic data. AI assisted with generating and refining target descriptions and reasoning text, screening candidate annotations, and cleaning and formatting training records. GPT-4o-mini and Qwen3-VL supported description generation and screening; offline teacher- and student-generated dialogue prefixes also contributed to multi-turn CoT data development. The geometry-guided trajectories, filtering procedures and stage-specific supervision are described in Appendices A and C.

Mathematical claims. AI assisted with developing proof arguments, deriving mathematical relationships and checking algebraic and geometric steps supporting the method's formulations, including spherical geometry and training objectives. The resulting definitions and derivations are presented in the method section and Appendix B.

Verification and responsibility. Checks on AI-assisted artifacts included code execution and tests, consistency checks on coordinates and data records, comparison of reported metrics with saved predictions and evaluation outputs, and consultation of original sources for references and benchmark protocols. The authors are responsible for reviewing the AI-assisted material and for the correctness, originality and integrity of the final manuscript, mathematical claims, code, data and reported results.

REFERENCES

Rahul Ahuja, Mudit Jain, Bala Murali Manoghar Sai Sudhakar, Venkatraman Narayanan, Pratik Likhar, Varun Ravi Kumar, and Senthil Yogamani. FishRoPE: Projective rotary position embeddings for omnidirectional visual perception. *arXiv preprint arXiv:2604.10391*, 2026. URL https://arxiv.org/abs/2604.10391.

Anthropic. Introducing Claude Sonnet 5. https://www.anthropic.com/news/claude-sonnet-5, 2026. Accessed September 25, 2026.

Iro Armeni, Sasha Sax, Amir R. Zamir, and Silvio Savarese. Joint 2d-3d-semantic data for indoor scene understanding. *ArXiv*, abs/1702.01105, 2017. URL https://arxiv.org/abs/1702.01105.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-VL technical report. *arXiv preprint arXiv:2511.21631*, 2025. URL https://arxiv.org/abs/2511.21631.

Yaniv Benny and Lior Wolf. Sphereuformer: A u-shaped transformer for spherical 360 perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 940–950, June 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Benny_SphereUFormer_A_U-Shaped_Transformer_for_Spherical_360_Perception_CVPR_2025_paper.html.

Vukasin Bozic, Isidora Slavkovic, Dominik Narnhofer, Nando Metzger, Denis Rozumny, Konrad Schindler, and Nikolai Kalischek. Unified panoramic geometry estimation via multi-view foundation models. *arXiv preprint arXiv:2605.26368*, 2026. URL https://arxiv.org/abs/2605.26368.

Angel X. Chang, Angela Dai, Thomas A. Funkhouser, Maciej Halber, Matthias Nießner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgb-d data in indoor environments. In *2017 International Conference on 3D Vision (3DV)*, pp. 667–676, 2017. URL https://arxiv.org/abs/1709.06158.

Volkan Cirik, Taylor Berg-Kirkpatrick, and Louis-Philippe Morency. Refer360°: A referring expression recognition dataset in 360° images. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7189–7202, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.644. URL https://aclanthology.org/2020.acl-main.644/.

Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1801.10130.

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

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. URL https://arxiv.org/abs/1901.05555.

Zihao Dongfang, Xu Zheng, Ziqiao Weng, Yuanhuiyi Lyu, Danda Pani Paudel, Luc van Gool, Kailun Yang, and Xuming Hu. Are multimodal large language models ready for omnidirectional spatial reasoning? *ArXiv*, abs/2505.11907, 2025. URL https://arxiv.org/abs/2505.11907.

Google DeepMind. Gemini 3.1 Pro Model Card. https://deepmind.google/models/model-cards/gemini-3-1-pro/, 2026. Accessed September 25, 2026.

Haocong He, Chenfei Liao, Zichen Wen, Zihao Dongfang, Xu Zheng, Bin Ren, Chang Su, Zixin Zhang, Harold Haodong Chen, Hongfei Zhang, Weijia Li, Kailun Yang, Conghui He, Xuming Hu, Nicu Sebe, and Linfeng Zhang. OmniCoT: A benchmark for global and multi-step panoramic reasoning. *arXiv preprint arXiv:2606.30378*, 2026. URL https://arxiv.org/abs/2606.30378.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. URL https://arxiv.org/abs/1705.07115.

Mukul Khanna, Ram Ramrakhya, Gunjan Chhablani, Sriram Yenamandra, Theophile Gervet, Matthew Chang, Zsolt Kira, Devendra Singh Chaplot, Dhruv Batra, and Roozbeh Mottaghi. GOAT-Bench: A benchmark for multi-modal lifelong navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16373–16383, 2024. URL https://arxiv.org/abs/2404.06609.

Kuula. Kuula: 360 virtual tours and panoramic images. https://kuula.co/, 2026. Accessed: 2026-05.

Zekai Lin and Xu Zheng. Panoenv: Exploring 3d spatial intelligence in panoramic environments with reinforcement learning. *arXiv preprint arXiv:2602.21992*, 2026. URL https://arxiv.org/abs/2602.21992.

Zhixin Ling, Zhen Xing, Xiangdong Zhou, Manliang Cao, and Guichun Zhou. Panoswin: A pano-style swin transformer for panorama understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17755–17764, June 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Ling_PanoSwin_A_Pano-Style_Swin_Transformer_for_Panorama_Understanding_CVPR_2023_paper.html.

Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, and Lei Zhang. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision – ECCV 2024*, pp. 38–55, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-72970-6. doi: 10.1007/978-3-031-72970-6_3. URL https://link.springer.com/chapter/10.1007/978-3-031-72970-6_3.

Yan Ma, Weiyu Zhang, Tianle Li, Linge Du, Xuyang Shen, and Pengfei Liu. What does vision tool-use reinforcement learning really learn? disentangling tool-induced and intrinsic effects for crop-and-zoom. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.01334.

OpenAI. Hello gpt-4o, 2024a. URL https://openai.com/index/hello-gpt-4o/.

OpenAI. GPT-4o mini: Advancing cost-efficient intelligence, July 2024b. URL https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/.

OpenAI. GPT-5.5 Model. https://developers.openai.com/api/docs/models/gpt-5.5, 2026a. Accessed September 25, 2026.

OpenAI. GPT-5.6 Sol Model. https://developers.openai.com/api/docs/models/gpt-5.6-sol, 2026b. Accessed September 25, 2026.

Ji Qi, Ming Ding, Weihan Wang, Yushi Bai, Qingsong Lv, Wenyi Hong, Bin Xu, Lei Hou, Juanzi Li, Yuxiao Dong, and Jie Tang. CogCoM: A visual language model with chain-of-manipulations reasoning. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Fg0eo2AkST.

Yuankai Qi, Qi Wu, Peter Anderson, Xin Wang, William Yang Wang, Chunhua Shen, and Anton van den Hengel. Reverie: Remote embodied visual referring expression in real indoor environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Qi_REVERIE_Remote_Embodied_Visual_Referring_Expression_in_Real_Indoor_Environments_CVPR_2020_paper.html.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026. URL https://qwen.ai/blog?id=qwen3.5.

Jiahui Ren, Mochu Xiang, Jiajun Zhu, and Yuchao Dai. Panosplatt3r: Leveraging perspective pretraining for generalized unposed wide-baseline panorama reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 28959–28969, October 2025. URL https://arxiv.org/abs/2507.21960.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. URL https://arxiv.org/abs/2402.03300.

Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. RoFormer: Enhanced transformer with rotary position embedding. *arXiv preprint arXiv:2104.09864*, 2021. URL https://arxiv.org/abs/2104.09864.

Cheng Sun, Chi-Wei Hsiao, Min Sun, and Hwann-Tzong Chen. HorizonNet: Learning room layout with 1d representation and pano stretch data augmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1047–1056, 2019. URL https://arxiv.org/abs/1901.03861.

Changpeng Wang, Xin Lin, Junhan Liu, Yuheng Liu, Zhen Wang, Donglian Qi, Yunfeng Yan, and Xi Chen. PanoWorld: Towards spatial supersensing in 360° panorama world, 2026. URL https://arxiv.org/abs/2605.13169.

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*, 2025a. URL https://arxiv.org/abs/2505.15966.

Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, et al. InternVL3.5: Advancing open-source multimodal models in versatility, reasoning, and efficiency. *arXiv preprint arXiv:2508.18265*, 2025b. URL https://arxiv.org/abs/2508.18265.

Mingyuan Wu, Jingcheng Yang, Jize Jiang, Meitang Li, Kaizhuo Yan, Hanchao Yu, Minjia Zhang, Chengxiang Zhai, and Klara Nahrstedt. VTool-R1: VLMs learn to think with images via reinforcement learning on multimodal tool use. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2505.19255.

Penghao Wu and Saining Xie. V*: Guided visual search as a core mechanism in multimodal LLMs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. URL https://arxiv.org/abs/2312.14135.

Liu Yang, Huiyu Duan, Ran Tao, Juntao Cheng, Sijing Wu, Yunhao Li, Jing Liu, Xiongkuo Min, and Guangtao Zhai. Odi-bench: Can mllms understand immersive omnidirectional environments? *ArXiv*, abs/2510.11549, 2025. URL https://arxiv.org/abs/2510.11549.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2210.03629.

Heyang Yu, Yinan Han, Xiangyu Zhang, Baiqiao Yin, Bowen Chang, Xiangyu Han, Xinhao Liu, Jing Zhang, Marco Pavone, Chen Feng, Saining Xie, and Yiming Li. Thinking in 360°: Humanoid visual search in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026. URL https://arxiv.org/abs/2511.20351.

Jingdong Zhang, Yizhou Wang, Zhengzhong Tu, Xin Li, Wenping Wang, and Xiaohang Zhan. Beyond thinking: Imagining in 360° for humanoid visual search. *arXiv preprint arXiv:2605.09146*, 2026. URL https://arxiv.org/abs/2605.09146.

Xinshen Zhang, Zhen Ye, and Xu Zheng. Towards omnidirectional reasoning with 360-r1: A dataset, benchmark, and grpo-based method. *arXiv preprint arXiv:2505.14197*, 2025. URL https://arxiv.org/abs/2505.14197.

Duo Zheng, Shijia Huang, Lin Zhao, Yiwu Zhong, and Liwei Wang. Towards learning a generalist model for embodied navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13624–13634, June 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Zheng_Towards_Learning_a_Generalist_Model_for_Embodied_Navigation_CVPR_2024_paper.html.

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. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2505.14362.

Yikang Zhou, Tao Zhang, Dizhe Zhang, Shunping Ji, Xiangtai Li, and Lu Qi. Dense360: Dense understanding from omnidirectional panoramas. *ArXiv*, abs/2506.14471, 2025. URL https://arxiv.org/abs/2506.14471.

Muzhi Zhu, Hao Zhong, Canyu Zhao, Zongze Du, Mingyu Liu, Zheng Huang, Anzhou Li, Hao Chen, Cheng Zou, Jingdong Chen, Ming Yang, and Chunhua Shen. ACTIVE-o3: Empowering MLLMs with active perception via pure reinforcement learning. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2505.21457.

A Dataset and Supervision

A.1 Sources and Splits

Table 5 summarizes the corpus from Matterport3D (Chang et al., 2017), 2D-3D-S (Armeni et al., 2017), Kuula (Kuula, 2026) and PanoInfinigen (Bozic et al., 2026). Training uses four yaw orientations per target; validation and test each use one scored orientation. Targets, source panoramas and augmented trajectories are counted separately. Real panoramas range from 2048×1024 to 8192×4096; PanoInfinigen images are 4096×2048.

Table 5: Dataset statistics before stage-specific replay. Panorama counts exclude augmentation variants.

Source / split	Targets	Source panoramas	Trajectories
Matterport3D (train)	2,602	2,119	10,408
2D-3D-S (train)	378	294	1,512
Kuula (train)	111	91	444
PanoInfinigen (train)	2,380	2,378	9,520
Train	5,471	4,882	21,884
Validation	319	319	319
Test	713	707	713

A.2 Annotation and Split Construction

GroundingDINO (Liu et al., 2025) proposes objects in cubemap views and verifies target-centered projections. GPT-4o-mini (OpenAI, 2024b) assists description generation and target-uniqueness checks for real panoramas; Qwen3-VL (Bai et al., 2025) assists PanoInfinigen screening. Human review checks label correctness, instance uniqueness, box completeness, clarity and distortion. Detector thresholds and source-license metadata accompany the annotations.

We remove training records that share source panoramas with the test split and assign PanoInfinigen targets by a deterministic panorama-level split. The resulting sets have disjoint recorded parent identifiers and checked PanoInfinigen image hashes. Training yaw offsets $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ consistently update images, target directions and camera azimuths while preserving elevation. Direction-balanced evaluation uses a separate frozen reorientation procedure (Appendix D.1).

A.3 Reference Trajectories and SFT Records

Geometry-guided reference trajectories cover coarse search, refinement and recovery, with 512×512 projected observations and templated reasoning. Target coordinates guide offline synthesis but never enter inference feedback. These reference trajectories differ from the actual SFT records: each optimization example supervises one current assistant reply with a frozen prefix, and later CoT stages also use dialogue prefixes generated offline by teacher and student models. Appendix C.1 reports the stage-specific sample counts, supervision and replay budgets.

B Geometric Construction and Supervision Details

We use the global angles and unit-ray map in Section 3.1: zero azimuth lies on the first Cartesian axis, positive azimuth points toward the second, and positive elevation (spherical latitude) toward the third. Trigonometric expressions use radians; tool arguments use degrees.

B.1 AtlasRoPE Frames and Phase Construction

Token rays and local frames. Let W_g, H_g be the merged visual-token grid dimensions, with $x_i \in \{0, \dots, W_g - 1\}$ and $y_i \in \{0, \dots, H_g - 1\}$. Pixel-center sampling gives

$$\theta_i = 2\pi \frac{x_i + 1/2}{W_g} - \pi, \qquad \phi_i = \pi \left(\frac{1}{2} - \frac{y_i + 1/2}{H_g} \right), \qquad \mathbf{d}_i = \mathbf{d}(\theta_i, \phi_i), \tag{5}$$

using the unit-ray function defined in Section 3.1. For chart center (α_k, β_k), the center, east and up unit vectors are

$$\begin{aligned}\mathbf{c}_k &= (\cos\beta_k \cos\alpha_k, \cos\beta_k \sin\alpha_k, \sin\beta_k),\\ \mathbf{e}_k &= (-\sin\alpha_k, \cos\alpha_k, 0),\\ \mathbf{u}_k &= (-\sin\beta_k \cos\alpha_k, -\sin\beta_k \sin\alpha_k, \cos\beta_k).\end{aligned} \tag{6}$$

Index k identifies a key/value group. Define local components $(c_{ki}, e_{ki}, u_{ki}) = (\mathbf{d}_i^\top \mathbf{c}_k, \mathbf{d}_i^\top \mathbf{e}_k, \mathbf{d}_i^\top \mathbf{u}_k)$. The chart coordinates are

$$\begin{aligned}\widetilde{\theta}_{ki} &= \operatorname{atan2}(e_{ki}, c_{ki}), & \widetilde{\phi}_{ki} &= \operatorname{atan2}\left(u_{ki}, \sqrt{c_{ki}^2 + e_{ki}^2}\right),\\ \widetilde{x}_{ki} &= a_k^x + \frac{W_g}{2\pi}\widetilde{\theta}_{ki}, & \widetilde{y}_{ki} &= a_k^y - \frac{H_g}{\pi}\widetilde{\phi}_{ki},\end{aligned} \tag{7}$$

where atan2 is the quadrant-aware arctangent and $a_k^x = W_g(\alpha_k/(2\pi)+1/2)-1/2$, $a_k^y = H_g(1/2-\beta_k/\pi)-1/2$ anchor the chart center in native grid units. This changes positional coordinates without resampling visual features.

Bounded cap and native fallback. Using $\mathrm{clip}_{[a,b]}(v) = \min(b, \max(a, v))$, we set

$$s_{ki} = \mathrm{clip}_{[0,1]}\left(\frac{c_{ki} - \cos r_{\mathrm{out}}}{\cos r_{\mathrm{in}} - \cos r_{\mathrm{out}}}\right), \qquad g_{ki} = s_{ki}^2(3 - 2s_{ki}), \tag{8}$$

with cap radii $r_{\mathrm{in}} = 60^\circ$ and $r_{\mathrm{out}} = 80^\circ$ converted to radians. The gate smoothly returns to native coordinates outside the inner cap and vanishes beyond the outer cap. Group $k = 0$ is fixed to $g_{0i} = 0$.

The fixed center list, in degrees and group order, is

$$\begin{gathered}(0,0),\ (0,90),\ (0,-90),\ (180,0),\\ (90,0),\ (-90,0),\ (45,45),\ (-135,-45).\end{gathered}$$

The backbone uses the first centers required by its key/value-group count. Equation 2 adds offsets to native phases in the height/width channels; each group's keys and associated query heads share these phases. Temporal channels and non-ERP positions remain native, and $\eta = 0$ recovers MRoPE exactly.

B.2 SSR Pool and Stage Sampling

The six replay strata use polar precedence: $\phi \geq 60^\circ$ is north and $\phi \leq -60^\circ$ south; otherwise $|\theta| \geq 150^\circ$ is a seam stratum and the rest central, each split by the sign of θ. The initial pool adds 469 records per stratum to 21,884 ordinary records, giving 24,698 examples. Equal quotas apply to the replay pool, not the combined distribution or each minibatch. Later scenario sampling can change these proportions; total stage counts appear in Table 6.

For a proper three-dimensional rotation R (an orthogonal matrix with determinant one), panorama reorientation and target transformation obey

$$I_R(\mathbf{d}) = I(R^{-1}\mathbf{d}), \qquad \mathbf{d}^{*,R} = R\mathbf{d}^*. \tag{9}$$

Here I maps a viewing ray to its panorama color, I_R is the reoriented panorama, and $\mathbf{d}^{*,R}$ is the rotated ground-truth direction. Camera parameters and coordinate-bearing supervision use the same transformed frame, and any tool views are rendered from the corresponding panorama. Only training sources are used. Controls substitute equal-count ordinary records matched by scenario, action and history length (Appendix C.1).

B.3 SDR Anchor, Normalization and Gradients

For example i, let $\mathcal{S}_i$ contain token indices supervised in the current assistant turn and let $a_i = \min \mathcal{S}_i - 1$. With decoder width D, the last-layer state $\mathbf{h}_{i,a_i} \in \mathbb{R}^D$ precedes the entire supervised response, including its CoT. A head with trainable matrix $A \in \mathbb{R}^{3\times D}$ and bias $\mathbf{b} \in \mathbb{R}^3$ computes

$$\mathbf{z}_i = A\mathbf{h}_{i,a_i} + \mathbf{b}, \qquad \widehat{\mathbf{d}}_i = \frac{\mathbf{z}_i}{\max(\|\mathbf{z}_i\|_2, \epsilon)}, \qquad \epsilon = 10^{-6}. \tag{10}$$

The stabilized vector has norm at most one. The target in Eq. 3 is the object's direction, not a coarse tool center. Scenario weights multiply both language and SDR losses; the auxiliary anchor excludes the current reply. For the first ten updates of each SDR stage, the anchor is detached while the head trains; afterward the auxiliary gradient reaches trainable backbone parameters. The head is discarded at inference. GRPO uses the same directional penalty with the state weights in Appendix C.3.

C Training and Interaction

C.1 Supervised Fine-Tuning

All EAGLE-360 and matched component models start from Qwen3-VL-8B-Instruct (Bai et al., 2025). Plain SFT disables AtlasRoPE, SSR and SDR and ends after multi-turn SFT; EAGLE-360-8b additionally receives interaction-focused SFT and GRPO. Matched component configurations keep their component flags throughout and use the fixed multi-turn checkpoint at update 574 without per-configuration selection. Each stage uses one epoch and seed 42; SSR-disabled controls replace replay with equal-count ordinary training records.

Table 6: SFT sample and optimization budgets. Replay is included in the example count.

Stage	Examples	Replay	Updates	Global batch
Coordinate SFT (matched controls)	24,698	2,814	6,175	4
Coordinate SFT (EAGLE-360-8b)	24,698	2,814	3,088	8
Single-turn CoT warm-up	24,698	2,814	3,088	8
Multi-turn SFT	4,589	1,561	574	8
Interaction-focused SFT	1,316	433	165	8

Optimization. CoT stages use global batch 8 on V100 GPUs. Low-rank adaptation (LoRA) (Hu et al., 2022) uses rank 16, alpha 32 and dropout 0.05, with cosine decay, zero weight decay and gradient clipping at 1.0. Single-turn CoT training uses learning rates $5\times10^{-6}/2.5\times10^{-7}/2.5\times10^{-6}$ for the language model, visual LoRA and visual-token merger, respectively, with 3% warm-up. The merger maps visual features to decoder inputs. Multi-turn stages freeze the vision encoder and visual-token merger, use a language learning rate 2×10^{-6} and 5% warm-up. SDR uses a head learning rate 2.5×10^{-4} and weight 0.02. Context is capped at 16,384 tokens, with 262,144–524,288 visual pixels; language computation uses 16-bit floating point (FP16), with 32-bit floating point (FP32) for the vision encoder and trainable LoRA parameters. Upstream coordinate-SFT budgets differ between the main model and matched controls as shown in Table 6; only the matched component study controls that budget.

Supervision. Single-turn CoT retains the first-turn reasoning structure but substitutes the precise ground-truth coordinates in the tool call. Later stages include direct answers, refinement, search/recovery and confirmation. Only the current assistant response—CoT, tool call or final answer, and end token—receives language loss; history tokens are masked. Multi-turn scenario weights for initial decision, refinement, search recovery, termination, student recovery and first-turn replay are $0.12/0.12/0.10/0.06/0.20/0.40$. Interaction-focused weights for entry, finish, follow-up, recovery, direct answer and replay are $0.25/0.25/0.10/0.10/0.15/0.15$. Both language and SDR losses use these weights. All component configurations share frozen offline prefixes. The component results are single-seed comparisons, and the reported stage was selected exploratorily after examining the later stage.

C.2 Interaction Format

The policy uses global azimuth/elevation and may answer directly or request a perspective view. Each reply contains nonempty CoT text followed by exactly one final answer or tool call:

Response schema

```
<think>visual observation and action justification</think>
<answer>azimuth: a, elevation: e</answer>
or
<tool_call>
{"name":"rotate_and_project_panorama",
 "arguments":{"azimuth":a,"elevation":e,
 "fov_degrees":f}}
</tool_call>
```

Here $a \in [-180, 180]$, $e \in [-90, 90]$ and $f \in [20, 120]$ denote azimuth, elevation and FoV in degrees, respectively; all must be finite. Tool feedback contains a 512×512 view and camera metadata, never ground-truth annotations. Greedy evaluation permits three tool calls, four assistant replies, 512 new tokens per reply and a 16,384-token context. Invalid outputs and exhausted budgets remain failures, without parser repair or camera-center fallback.

C.3 GRPO Training and Checkpoint Selection

Initialization and sampling. GRPO starts from the AtlasRoPE+SSR+SDR interaction-focused SFT model. Its adapter is merged and frozen; a new language-only LoRA adapter (rank 16, alpha 32, zero dropout) and the SDR head are trained. The vision encoder and visual-token merger remain frozen. The selected training run processes 1,280 training questions, with held-out parent/image checks; teacher responses do not enter on-policy rollouts.

Two RTX PRO 6000 GPUs process eight questions per update. Each question produces four first responses (roots); each executed first tool branches into four continuations. Sampling uses temperature 0.7, top-$p = 1$ and the interaction limits in Appendix C.2. All invalid trajectories are retained. AdamW uses policy/head rates $5 \times 10^{-6}/2.5 \times 10^{-4}$, $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$ and zero weight decay; gradients are clipped at 1.0. One synchronized optimizer update follows each rollout batch. Rollout/replay use FP16 autocast with FP32 trainable parameters.

Branch sampling and advantages. For each question q, index the four first responses (roots) by $j = 1, \ldots, 4$ and each root's terminal continuations (leaves) by $k = 1, \ldots, K_j$. A terminating root has $K_j = 1$; an executed first tool has $K_j = 4$. For the reward equations, express the great-circle error d_{gc} and acceptance radius $r = \delta_{\text{bbox}}/2$ in degrees. A valid final direction has hit indicator $h = \mathbf{1}[d_{\text{gc}} \le r]$, where $\mathbf{1}[\cdot]$ is the indicator function. The terminal return is

$$R_{\text{final}} = h + \frac{0.30}{1 + (d_{\text{gc}}/\max(r, 10^{-3}))^2} + 0.05, \tag{11}$$

and is -0.25 when no valid final answer is produced. A *correctness-gated, bounded* bonus encourages tool adoption. Let z_{qj} indicate that root j executes a tool, $K_j \in \{1, 4\}$ its leaf count, and $p_q = \frac{1}{4}\sum_{j=1}^{4} z_{qj}$. For leaf k,

$$\begin{aligned} b_{qjk} &= 0.4(1 - p_q) z_{qj} \mathbf{1}[\text{valid final answer and hit}], \\ \bar{R}_{qj} &= \frac{1}{K_j} \sum_{k=1}^{K_j} (R_{\text{final},qjk} + b_{qjk}), \\ A^{\text{root}}_{qj} &= \bar{R}_{qj} - \frac{1}{4} \sum_{\ell=1}^{4} \bar{R}_{q\ell}. \end{aligned} \tag{12}$$

The adoption rate counts unique roots. The bonus does not grow with later calls; root advantages are mean-centered without variance normalization.

For a tool root with first camera-center great-circle error d_c and hit indicator $h_c = \mathbf{1}[d_c \le r]$, each valid continuation receives

$$A^{\text{cont}} = (h - h_c) + 0.5 \tanh((d_c - d_{\text{gc}})/5)\,, \tag{13}$$

or -1.5 for an invalid final output. This advantage applies to continuation tokens without centering or addition to the root return. The camera center is an action parameter, not an independently submitted answer.

Token objective and auxiliary supervision. Let $\mathcal{S}_q$ contain each root response once and every generated response in its continuations. Segment s has advantage A_s from Eqs. 12–13, weight $w_s = 1$ for a root and $w_s = 1/K_j$ for a continuation of root j, and token ratio $\rho_{st} = \pi_\vartheta(y_{st} \mid x_{st})/\pi_{\text{old}}(y_{st} \mid x_{st})$. Here t indexes generated tokens, y_{st} is a token, x_{st} is its preceding multimodal context, and π_ϑ, π_{old} and π_{ref} denote the current, rollout and frozen SFT policies; ϑ denotes policy parameters. The advantage A_s is constant across tokens in a segment. For a global question batch of size $M = 8$, the minimized loss is

$$\begin{aligned}
\mathcal{L}_{\text{policy}} &= \frac{1}{4M \cdot 128} \sum_{q=1}^{M} \sum_{s \in \mathcal{S}_q} w_s \sum_{t \in s} [\ell_{st} + 0.05\, k_{st}]\,, \\
\ell_{st} &= -\min\{\rho_{st} A_s, \text{clip}_{[0.8,1.2]}(\rho_{st}) A_s\}, \\
k_{st} &= e^{\Delta_{st}} - \Delta_{st} - 1, \qquad \Delta_{st} = \log \pi_{\text{ref}}(y_{st} \mid x_{st}) - \log \pi_\vartheta(y_{st} \mid x_{st}).
\end{aligned} \tag{14}$$

Here ℓ_{st} is the clipped policy loss and k_{st} is the per-token KL penalty estimate. Rollout, current and reference policy likelihoods use the same temperature-0.7 distributions. Only generated assistant tokens enter this sum. The fixed 128 is a loss normalizer, not a generation-length limit. The additional loss is $0.02\mathcal{L}_{\text{SDR}}$, using the per-state direction loss in Eq. 3 with the state weights specified next rather than the SFT scenario weights. If any root uses a tool, half the per-question auxiliary weight is placed at the initial pre-response state and half is distributed equally over the unique first post-tool states; otherwise all auxiliary weight remains at the root. The current assistant reply is excluded from each auxiliary anchor.

Checkpoint selection. Evaluation on all 319 validation questions occurs every 32 updates. Eligible checkpoints exceed a fixed SFT reference evaluated on the same hardware in correct-prediction count and number of questions with model-initiated tool use while reducing full-set GCD; they are ranked by correct-prediction count, then GCD, then the number of questions with tool use. The rule selects cumulative update 160, frozen before evaluation on the 713-question test set and H*Bench evaluation.

D Evaluation Protocols

D.1 Direction-Balanced Evaluation

The frozen test contains 713 questions: Front/Right/Back/Left/Up each have 119 and Down has 118. A fixed-seed procedure uses annotations, not model predictions, to reassign the minimum necessary targets: 235 question-associated views are reoriented and 478 retain their original orientation. Seven direction-dependent queries keep their original orientation. Validation has 319 separate questions.

Reorientation applies Eq. 9 to the full ERP and target, preserving resolution and angular radius. Reassigned polar targets have uniform azimuth and $|\phi| \in [60^\circ, 80^\circ]$; horizontal targets lie within 35° of a band's central azimuth with $\phi \in [-30^\circ, 30^\circ]$. Elevation is sampled uniformly in $\sin\phi$ within the applicable interval. Evaluation assigns Up/Down at $\phi \ge 45^\circ/\phi \le -45^\circ$; remaining targets use four 90° azimuth bands centered at $0^\circ, 90^\circ, 180^\circ, -90^\circ$. Models receive no target, direction-band label or rotation metadata.

Models and input budgets. The comparison includes InternVL3.5-4B/8B (Wang et al., 2025b), Qwen3-VL-4B/8B (Bai et al., 2025), Qwen3.5-4B/9B (Qwen Team, 2026), GPT-4o (OpenAI, 2024a), Gemini 2.5 Pro (Comanici et al., 2025), Claude Sonnet 5 (Anthropic, 2026), GPT-5.5 (OpenAI, 2026a), Gemini 3.1 Pro Preview (Google DeepMind, 2026), GPT-5.6 Sol (OpenAI, 2026b), public PanoWorld (Wang et al., 2026), Hstar (Yu et al., 2026), Plain SFT and EAGLE-360. General

open-weight models use FP16 greedy decoding and 512 output tokens; InternVL allows 12 tiles plus a thumbnail, while Qwen uses up to 524,288 pixels. Proprietary models use high-detail input and model-specific decoding budgets. Direct-coordinate prompts do not request explicit CoT; provider-internal reasoning is marked in Table 1.

Public PanoWorld uses 2,048 visual tokens and a 512-token cap without EAGLE-360-specific fine-tuning. Hstar uses the released HVS-3B policy and native local-view search: an initial 100° FoV at yaw/pitch $0^\circ/0^\circ$, up to ten actions and 1,024 tokens per reply. Plain SFT and EAGLE-360-8b start with the full ERP and use Appendix C.2. Native image processing and tool budgets are retained for each model.

Scoring and feedback controls. Every configuration is scored on all 713 question IDs with the same targets and BFoV criterion (Eq. 1). Geometric correctness is separate from answer-only formatting; missing or invalid coordinates count as incorrect with 180° angular error. Interactive models require an explicit valid terminal answer.

The six feedback conditions in Table 4 share one evaluation environment with eight NVIDIA V100 GPUs and a fixed policy. Real-view and mismatched-view conditions share the first response and camera request; a mismatched view comes from a different panorama. Low-resolution initialization uses 512×256 images, while tool views remain 512×512 projections from full-resolution sources. The no-tool condition adds an instruction disabling tools; all conditions retain CoT. Each condition includes all 713 questions. These results are reported separately from the main evaluation on NVIDIA RTX PRO 6000 Blackwell GPUs.

D.2 H*Bench Transfer

H*Bench has 600 object-search and 400 path-search questions, each with four start headings. Under full-ERP initialization the headings give identical inputs, so 1,000 actual rollouts produce 4,000 episode weights, not independent samples. Plain SFT and EAGLE-360-8b use the same images, coordinate adapter, scorer and interaction budget, without benchmark-specific training or checkpoint selection. Labels enter scoring only.

Table 2 uses a fixed version of the official scorer; its seam-handling branch omits the object-search elevation check. We also audited a scorer that restores this check. HOS uses the official angular success window, whereas HPS evaluates azimuth; their angular errors are not pooled. Hstar literature values come from Table 1 of Yu et al. (2026); dagger-marked Overall entries use $0.6\,\mathrm{HOS} + 0.4\,\mathrm{HPS}$ from rounded task rates. PanoWorld's values come from Table 3(a) of Wang et al. (2026), retaining its published Overall. Literature rows preserve their training and initial-observation conditions, which differ from our shared ERP setting.

D.3 Code, Model, and Test Data

The evaluation code is available through the anonymous repository:

`https://anonymous.4open.science/r/EAGLE360Review2027/`

The GRPO checkpoint and the 713-question direction-balanced test set are provided in the anonymous model archive and anonymous test-data archive, respectively. A downloadable code archive is also available. The repository README provides the environment setup, local asset paths, and commands for inference and scoring.

E Qualitative Examples

Figures 3–6 present four selected two-turn trajectories from a separate qualitative rerun of the validation-selected EAGLE-360-8b policy on the direction-balanced test set. Each case shows the panorama, model outputs, returned tool view, and final result. GT, Pred and Q denote the ground truth, prediction and query; HIT indicates a correct prediction under Eq. 1. These examples illustrate interaction; aggregate results are reported in Section 5.

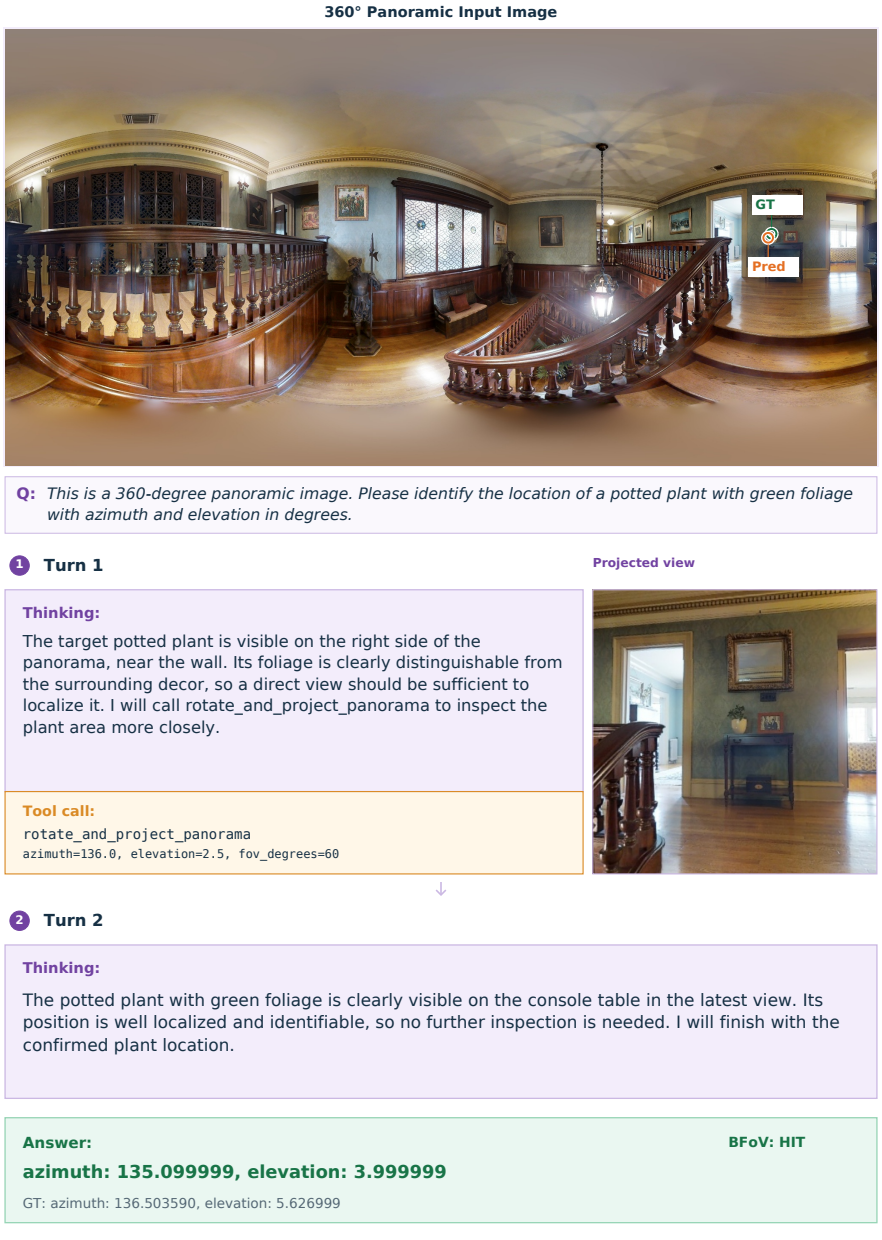


Figure 3: An inference case of EAGLE-360: a potted plant.

360° Panoramic Input Image



Q: *This is a 360-degree panoramic image. Where is a toilet located? Please provide the spherical coordinates (azimuth and elevation in degrees).*

1 Turn 1

Thinking:

The toilet is clearly visible on the right side of the panorama, with its bowl and tank identifiable. A closer view would help confirm the exact position and orientation, so I will inspect the toilet area with a targeted perspective.

Tool call:

```
rotate_and_project_panorama
azimuth=153.999999, elevation=-7.7, fov_degrees=60
```

Projected view



2 Turn 2

Thinking:

The toilet is clearly identifiable in the latest view, with its bowl, seat, and tank visible. The target is localized enough to finish with the answer.

Answer: **BFoV: HIT**

azimuth: 154.999999, elevation: -7.7

GT: azimuth: 155.230182, elevation: -7.702547

Figure 4: An inference case of EAGLE-360: a toilet.

360° Panoramic Input Image



Q: *This is a 360-degree panoramic image. I need to locate a plant. Please give me the azimuth and elevation in degrees.*

1 Turn 1

Thinking:

The plant is visible as a small dark object near the hallway wall, so I would inspect it with a local view to confirm and localize.

Tool call:

```
rotate_and_project_panorama
azimuth=57.15656, elevation=-2.75, fov_degrees=60
```

Projected view



2 Turn 2

Thinking:

The latest view clearly shows the plant in a dark pot on the floor near the hallway wall. The plant is identifiable and localized enough to finish.

Answer: **BFoV: HIT**

azimuth: 55.38977, elevation: -2.35977

GT: azimuth: 56.111181, elevation: -2.259808

Figure 5: An inference case of EAGLE-360: a plant near a hallway.

360° Panoramic Input Image



Q: *This is a 360-degree panoramic image. Can you find a green glass bottle and determine its angular position in degrees?*

1 Turn 1

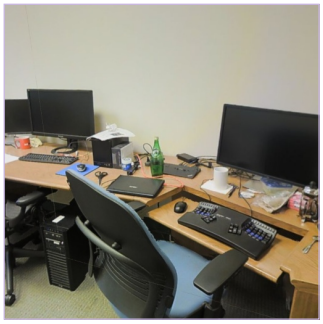
Thinking:

The green glass bottle is clearly visible on the right side of the panorama, sitting on the desk near the computer monitors. It is identifiable by its distinct green color and cylindrical shape. A closer view would help confirm its exact location, so I will inspect the right desk area with a targeted perspective.

Tool call:

```
rotate_and_project_panorama
azimuth=158.75, elevation=-18.7, fov_degrees=60
```

Projected view



2 Turn 2

Thinking:

The green glass bottle is clearly visible on the right side of the desk in the latest view, sitting between the monitors and a keyboard. Its green color and cylindrical shape are unmistakable. The target is localized and sufficiently confirmed, so no further inspection is needed.

Answer: **BFoV: HIT**

azimuth: 157.177, elevation: -18.7

GT: azimuth: 157.538361, elevation: -17.431349

Figure 6: An inference case of EAGLE-360: a green glass bottle.