Z-RoboSTAR: Building and Benchmarking Robotic System 2 Agents for Reasoning in Sustained Long-Horizon Manipulation

Anonymous authors
Paper under double-blind review

Abstract

Sustainable long-horizon robotic manipulation requires a deliberative System 2 agent that can reason, plan, and sustain progress toward a goal as the world evolves. Complex multi-step manipulation changes the world state, while interventions may invalidate prior beliefs, requiring System 2 to continually track progress from partial observations. We present Z-RoboSTAR, a harness that enables vision-language models (VLMs) to act as System 2 agents, centered on a dynamically updated symbolic scene graph that serves as a persistent shared-state interface to synchronize scene understanding, task progress, and planning, for failure recovery and resolving reconciliation. To efficiently compare open- and closed-source VLMs for our System 2 harness, we construct Z-RoboSTAR-Bench, an offline open-loop benchmark and evaluation engine spanning 57 tabletop scenes with diverse spatiotemporal dependencies and interventions. The model ranking is preserved under human-executed closed-loop evaluation, supporting the benchmark's relevance to real-world agent performance. Beyond evaluation, we develop Z-RoboSTAR-27B through trajectory-level distillation to improve its robustness over prolonged execution. When deployed on a Galaxea R1 Lite with a fine-tuned $\pi_{0.5}$ System 1 controller, Z-RoboSTAR-27B achieves a 66.9% task success rate on our real-world long-horizon manipulation tasks, ranking second only to GPT-5.5. These results demonstrate effective System 2 coordination of complex, long-horizon manipulation in real-world scenes.

1 Introduction

In complex long-horizon robotic manipulation, dual-system architectures that couple System 2 reasoning with System 1 execution are attracting attention (Zhou et al., 2026; Liu et al., 2026a; Shi et al., 2025). System 2 interprets task goals, selects subtasks based on the current state and progress, and revises decisions using feedback (Huang et al., 2022; Rana et al., 2023; Huang et al., 2024; Liu et al., 2026a; Wang et al., 2026; Duan et al., 2024; Shi et al., 2025; Li et al., 2026), while System 1 executes these commands through low-level policies, including vision-language-action (VLA) models (Bjorck et al., 2025; Intelligence et al., 2025; Liu et al., 2026b). Recent advances in vision-language models (VLMs) (OpenAI, 2026; Qwen Team, 2025; Wang et al., 2025; Comanici et al., 2025) provide a foundation for the System 2 component: their semantic knowledge, visual grounding, and reasoning capabilities support instruction understanding and scene interpretation for robotic manipulation (Driess et al., 2023; Jiang et al., 2023; Brohan et al., 2023; Huang et al., 2023). For long-horizon manipulation, System 2 must do more than propose an initial plan: *it must sustain progress toward the goal throughout execution as the world undergoes complex state transitions*.

Existing work has improved System 2 planning and execution monitoring through affordance grounding (Ahn et al., 2022), structured scene representations (Rana et al., 2023; Zhou et al., 2026; Wang et al., 2026), environmental feedback (Huang et al., 2022; Liu et al., 2026a; Shi et al., 2025), explicit state memory (Zhou et al., 2026; Wang et al., 2026), and failure-driven replanning (Huang et al., 2022; Duan et al., 2024; Zhou et al., 2026; Li et al., 2026). These mechanisms help agents generate feasible actions, incorporate execution feedback, and recover from individual failures. Long-horizon manipulation, however, *requires these capabilities to operate over a shared, continually updated representation*, so that scene changes are consistently reflected in both the agent's scene belief

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.

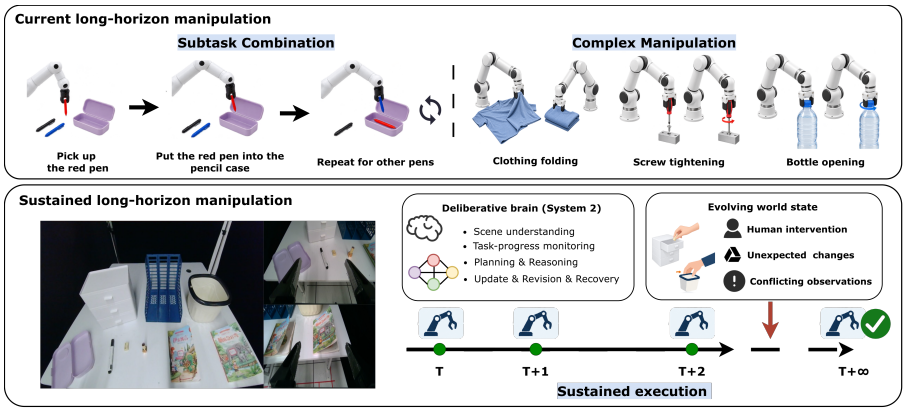


Figure 1: **Our focus: long-horizon manipulation under evolving world states.** The top panel illustrates subtask composition and complex manipulation skills, such as clothing folding, screw tightening, and bottle opening. We study how System 2 sustains goal-directed execution as robot actions, human interventions, and conflicting observations require it to update its estimate of the physical scene and reassess task progress. The bottom-left panel shows our tabletop setting: given a goal such as organizing the desk, System 2 must guide System 1 execution over extended interactions, including recovery when scene changes invalidate previously completed subtasks.

(i.e., estimate of object states and spatial relations) and its estimate of task progress. During System 1 execution, robot actions may unexpectedly alter multiple object relations, human interventions may invalidate completed subtasks, and new observations may conflict with the maintained scene belief. Consequently, successfully executing the latest action does not necessarily imply continued progress toward the overall goal when other task-relevant conditions change during execution.

This creates a coupled state-maintenance problem. After each interaction, System 2 must determine what changed, reconcile new visual evidence with its prior scene belief, and reassess which task conditions remain satisfied. Consider a robot that correctly places an object, which is later moved by another action or human intervention. If System 2 updates only the object's location, the task monitor may still mark the corresponding subtask as complete. If it checks only the latest System 1 action, it may miss that a previously satisfied goal condition has been invalidated. *System 2 must therefore update its scene belief and task progress together.*

We present Z-RoboSTAR, an agentic System 2 harness for VLMs centered on *a dynamically updated symbolic scene graph*. The structured graph represents objects as nodes and their relations as edges, providing *a shared state interface* for scene understanding, task-progress monitoring, and planning. Rather than treating the graph as static context, Z-RoboSTAR updates its nodes and relations throughout execution. After each interaction, action-outcome evaluation and global visual reconciliation update the graph, capturing changes beyond the action's expected effects. The revised state informs task progress and planning. Operationally, the graph is both the shared *read interface* for task monitoring, planning, and action validation, and the *write target* of action-outcome updates and global visual reconciliation. Accordingly, the System 2 modules share the revised scene state.

Identifying which VLMs perform best within our System 2 harness requires efficient, controlled comparisons. We therefore construct Z-RoboSTAR-Bench, an offline open-loop benchmark and evaluation engine. Its compiler generates executable tasks across 57 tabletop scenes and six task families with evolving spatiotemporal dependencies. Each nominal task has six intervention variants testing state tracking, progress monitoring, and recovery. Symbolic transitions and pre-collected observations avoid live physical execution, while explicit goal conditions provide task-level scoring.

We evaluate closed- and open-source VLMs within Z-RoboSTAR under nominal execution and controlled interventions. Offline open-loop and human-executed closed-loop evaluations yield *exactly the same model ordering in both task success and normalized goal progress* across the four tested models, despite differences in absolute performance. Ablations examine scene graphs, visual reconciliation, and task-progress monitoring. We build Z-RoboSTAR-27B, a 27B System 2 model for state tracking and feedback-driven embodied reasoning. On a Galaxea R1 Lite, both this model and GPT-5.5, operating within Z-RoboSTAR with a System 1 $\pi_{0.5}$ VLA controller, can complete complex, long-horizon manipulation tasks. We provide a demonstration in the supplementary material.

Our contributions are threefold:

- We introduce Z-RoboSTAR, a System 2 harness that synchronizes scene estimates, task progress, and planning through a dynamically updated symbolic scene graph, and build the efficient Z-RoboSTAR-27B model for embodied planning.
- We construct Z-RoboSTAR-Bench, an offline open-loop benchmark and evaluation engine with 57 tabletop scenes, enabling comparison of open- and closed-source VLMs.
- We compare System 2 models, show matching four-model rankings across open-loop and human-executed closed-loop evaluation, and demonstrate real-world long-horizon manipulation using our 27B model and GPT-5.5 with a fixed System 1 VLA controller.

2 Related Work

2.1 Embodied Foundation Models

Foundation models have increasingly been adapted to embodied perception, reasoning, and acting. PaLM-E (Driess et al., 2023) integrates continuous sensory observations into language-model reasoning, supporting visual question answering, scene understanding, and textual high-level robot planning. Vision-language-action (VLA) models and generalist robot policies instead map multimodal observations and instructions directly to robot actions, such as RT-2 (Brohan et al., 2023), OpenVLA (Kim et al., 2024), $\pi_{0.5}$ (Intelligence et al., 2025), GR00T N1 (Bjorck et al., 2025), and Galaxea G0.5 (Liu et al., 2026b). The vision-language and action-generation modules are tightly coupled and jointly trained within a single VLA. In contrast, explicitly hierarchical systems such as Galaxea G0 (Liu et al., 2026b), Hi Robot (Shi et al., 2025), and HAMSTER (Li et al., 2025) use a high-level VLM to generate subtask instructions for a separately trained low-level policy. Recent embodied VLMs further target spatial reasoning, task planning, and execution understanding (Dang et al., 2026; Tan et al., 2026; Azzolini et al., 2025; Team et al., 2025). These works provide increasingly strong embodied competence. Our work complements this by studying how such capabilities are maintained and updated throughout extended interaction.

2.2 Embodied Agent Systems

Embodied agents as System 2 connect high-level reasoning to physical execution in System 1 through grounding, structured representations, and feedback. SayCan (Ahn et al., 2022) grounds language decisions in executable skills, while Inner Monologue (Huang et al., 2022) incorporates execution and scene feedback into subsequent decisions. SayPlan (Rana et al., 2023), VoxPoser (Huang et al., 2023), and ReKep (Huang et al., 2024) introduce structured spatial or constraint-based interfaces for grounded task planning and reactive manipulation in physical environments. Recent agent frameworks further integrate persistent memory, execution monitoring, skill orchestration, and recovery (Liu et al., 2026a; Zhou et al., 2026; Li et al., 2026). These works establish the importance of closing the loop between reasoning and execution. Our work also builds on this perspective by focusing on the state carried through that loop; that is, how scene beliefs and task progress are persistently maintained, updated, and revised.

2.3 Long-Horizon Robot Learning and Evaluation

Long-horizon robot learning requires policies to coordinate skills over extended execution. SPIRE (Zhou et al., 2024) combines task-and-motion planning with imitation and reinforcement learning, while Long-VLA (Fan et al., 2025) and PALM (Liu et al., 2026c) improve long-horizon VLA policies through phase modeling, affordance reasoning, and progress estimation. Benchmarks such as CALVIN (Mees et al., 2022), LIBERO (Liu et al., 2023), RoboCasa (Nasiriany et al., 2024), and VLABench (Zhang et al., 2025b) evaluate skill composition, knowledge transfer, and multi-step manipulation. More recent suites, including VLA-Arena (Zhang et al., 2025a) and RoboDojo (Chen et al., 2026), introduce controlled perturbations and broader sim-to-real evaluation. Complementing these benchmarks, our evaluation focuses on whether an embodied System 2 agent can consistently revise its scene belief and task progress under complex spatiotemporal dependencies, unexpected action-induced changes, human interventions, and conflicting observations.

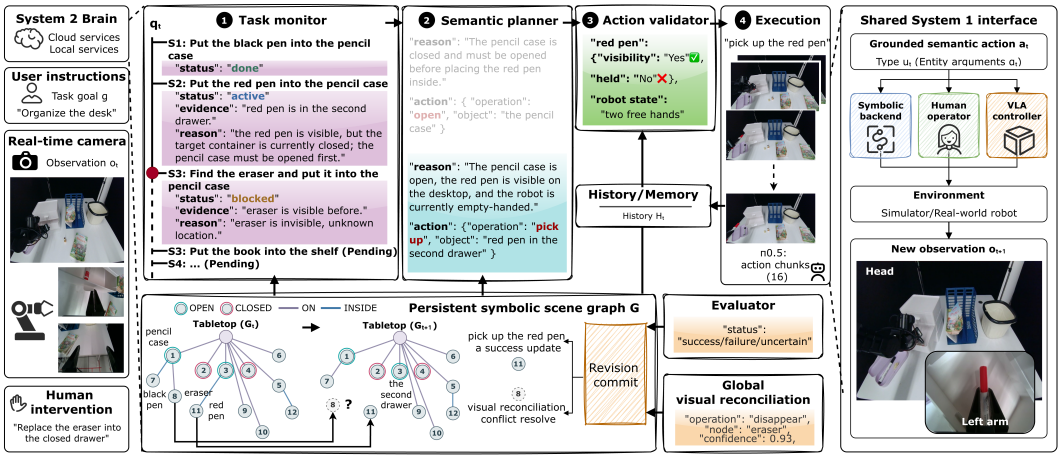


Figure 2: **Overview of** Z-RoboSTAR**.** The persistent symbolic scene graph provides shared state for task monitoring, semantic planning, and action validation. Action-outcome evaluation and global visual reconciliation update the graph from execution feedback. The revised state updates task progress and guides the next action planning. In the illustrated example, a human moves a previously visible eraser into a closed drawer. The task monitor marks the eraser subtask as `blocked` because its current location is unknown to the agent, while the red-pen subtask remains `active`. The planner selects grasping the visible pen (`pick up the red pen`) based on the current scene.

3 METHOD

3.1 PROBLEM FORMULATION AND SYSTEM OVERVIEW

We formulate long-horizon manipulation as a partially observed interactive process. As illustrated in Figure 2, at step t, the latent physical state s_t produces a visual observation $o_t \sim \mathcal{O}(s_t)$, and the agent receives a natural-language goal g. The world may change through the robot's actions or exogenous events, such as a human moving an object. Because the agent never observes latent s_t directly, neither the latest image nor the nominal effect of the latest action is sufficient to maintain an accurate world-state belief for subsequent planning. Z-RoboSTAR therefore maintains a symbolic scene belief G_t as a structured graph, a task-progress state q_t, and a bounded interaction history H_t. G_t records task-relevant objects, relations, and mutable attributes supported by accumulated evidence. q_t tracks task execution, progress, and evidence of completion. Together with the goal g, latest observation o_t, and preceding action outcome z_{t-1}, these states form the context

$$x_t = (g, o_t, G_t, q_t, H_t, z_{t-1}).$$

Given x_t, System 2 agent emits either a semantic action $a_t = (u_t, \alpha_t) \in \mathcal{A}_{\text{sem}}$ or a stop decision to System 1 action executor. Here, $\mathcal{A}_{\text{sem}}$ denotes the set of skill-level semantic commands supported by the executor, u_t specifies the action type, and α_t contains its grounded arguments. The executor, i.e., System 1, implemented as a symbolic backend, human operator, or VLA controller, then carries out the command and returns the next observation. As shown in Figure 2, Z-RoboSTAR operates in a closed loop. At each step, it uses G_t and q_t to select and validate the next action. After the executor returns a new observation o_{t+1}, the action-outcome evaluator assesses whether the intended effect occurred, while the scene reconciler detects additional or conflicting changes. Supported revisions are committed to G_{t+1}, and the task-progress state q_{t+1} is recomputed from the revised scene belief before the next action is planned.

3.2 SYMBOLIC SCENE GRAPH AS PERSISTENT SCENE STATE

We represent the persistent scene belief as a structured graph $G_t = (V_t, E_t, X_t)$. Each node $v_i \in V_t$ has a stable entity identifier and semantic description obtained either from prior knowledge or from a visual perception model (Liu et al., 2024; Neau, 2026). Edges encode task-relevant relations, including location, containment, part–whole structure, robot–object interactions, and occlusion. X_t stores mutable attributes such as visibility, open/closed state, held/released state, and the number of

arms required for manipulation. Each mutable fact is associated with evidence supporting its value, allowing the same representation to be read by task monitoring, planning, and action validation.

G_t is a revisable belief over the latent physical state s_t, constructed from accumulated evidence rather than direct access to the true state. It is initialized with task-relevant entities, relations, and attributes supported by available observations or prior knowledge. For example, an unseen pencil inside a closed drawer is not represented as a node unless its existence is supported by subsequent exploration. As new observations o_t arrive, visible facts are updated from current evidence, while previously observed entities that are no longer visible remain in G_t with their last supported state. If new evidence invalidates a previous belief without establishing a replacement, the corresponding fact is marked as unknown. Thus, nonvisibility alone does not imply relocation, nor does an action outcome imply that unobserved parts of the scene remain unchanged. Unknown facts can therefore prompt targeted information exploration when required for planning or validation.

Updates to G_t are expressed as typed edits over node attributes and relations. Candidate revisions may be supported by the preceding action outcome z_{t-1} from the evaluator and the latest observation o_t from global visual reconciliation with other accumulated evidence. Before an edit is committed, the system verifies entity identifiers, operation-specific fields, relation types, and structural consistency. This controlled write boundary allows the scene belief to incorporate supported changes while preventing unsupported assumptions about the latent state s_t.

3.3 Closed-Loop Harness Modules

Task monitoring and planning. The task monitor estimates progress from the maintained belief,

$$q_t = M_\theta(g, G_t, H_t),$$

and the planner selects one semantic action $a_t = \Pi_\theta(g, o_t, G_t, q_t, H_t)$. When a required fact is unknown, the action may instead inspect a task-relevant region or resolve some occlusion. The system stops only when current evidence supports every goal condition or exceeds the maximum steps.

Scene-conditioned action validation. Before System 1 execution, the validator checks the proposed action against its schema and the current graph,

$$d_t = V(a_t, G_t), \qquad d_t \in \{\text{accept}, \text{regenerate}\}.$$

It verifies arguments, entity references, and graph-grounded preconditions such as visibility, container state, and held objects. Rejected actions receive feedback for regeneration. Acceptance certifies admissibility under the current belief, not physical success.

Shared semantic System 1 interface. An accepted action is dispatched through a common executor adapter. The adapter allows different execution backends, such as symbolic transitions, a human operator, and a VLA controller in both the simulator and the real-world robot, and returns o_{t+1}.

Action-outcome evaluation. The execution evaluator determines whether the requested outcome occurred:

$$z_t = E_\phi(o_t, o_{t+1}, a_t), \qquad z_t \in \{\text{success}, \text{failure}, \text{uncertain}\}.$$

The result gates the action-conditioned update $\widetilde{G}_{t+1} = U(G_t, a_t, z_t)$. Success supports the expected effect, failure blocks it with feedback, and uncertainty leaves the relevant facts unresolved pending further evidence.

Global visual reconciliation. To capture changes beyond the requested effect, the reconciler compares the new observation with the action-conditioned belief,

$$\Delta_t = R_\phi(o_t, o_{t+1}, \widetilde{G}_{t+1}, a_t, z_t).$$

Δ_t contains evidence-supported edits to known entities, relations, and mutable attributes. This step can detect displaced objects, newly revealed entities, introduced occlusions, or observations that contradict earlier beliefs.

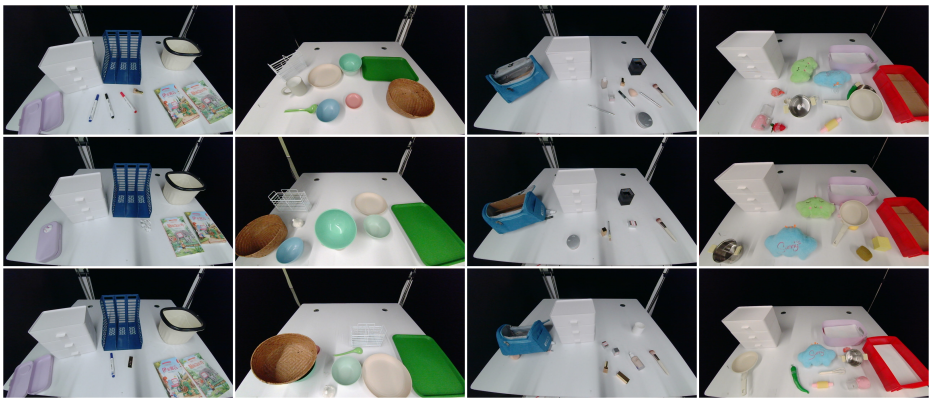


Figure 3: Visualization of representative tabletop scenes for sustained long-horizon tasks.

Revision and replanning. The commit step validates and atomically applies the proposed edits,

$$G_{t+1} = C(\widetilde{G}_{t+1}, \Delta_t), \qquad q_{t+1} = M_\theta(g, G_{t+1}, H_{t+1}).$$

Recomputing q_{t+1} from the revised graph keeps scene belief and task progress synchronized: a contradicted relation can reopen a completed subtask, a newly revealed entity can unblock a pending one, and an unresolved outcome can trigger inspection or recovery.

4 Z-RoboSTAR-Bench: Compiling Long-Horizon Benchmark with Evolving State

We focus on sustained long-horizon manipulation in the real world, where a robot must continue operating over extended interactions while the world may change in diverse and unexpected ways. Evaluating such capability requires more than increasing action-sequence length and sub-task numbers (Chen et al., 2026). We therefore establish a task-generation compiler to evaluate the closed-loop state-maintenance mechanisms above. It constructs executable long-horizon tasks whose difficulty arises from dependencies over an evolving world state, beyond action-sequence length. We use 57 tabletop scenes in our evaluation with 6 task families, including everyday desktop organization.

Task compilation. A compiled instance is defined as

$$\mathcal{I} = (G_0^\star, g, \mathcal{A}, \mathcal{O}, \Lambda),$$

where $G_t^\star$ denotes the latent benchmark scene graph containing diverse objects, attributes, relations, and spatial and temporal dependencies, g specifies the goal predicate decomposing the final goal, $\mathcal{A}$ defines the semantic action schemas, $\mathcal{O}$ governs partial observations, and Λ denotes the set of exogenous events, such as human intervention.

Spatial and temporal dependencies. Spatial dependencies place task-relevant facts outside the current observation, such as inside closed containers or behind occluders, requiring active exploration before they can inform later decisions. Temporal dependencies arise when later steps depend on the completion or information acquired from earlier steps, creating causal chains across the task. We quantify the difficulty of these evolving spatial states by the length of temporal causal chains and the number and depth of spatial occlusions that must be resolved.

Dynamic interventions. An intervention is represented as

$$\lambda = \langle \psi_\lambda, \delta_\lambda, \kappa_\lambda \rangle \in \Lambda,$$

comprising a semantic trigger, a graph rewrite, and an execution condition. Triggers depend on task progress rather than wall-clock time, preserving reproducible causal structure. Instantiated events include action failure, state regression, rollback of completed subgoals, unexpected relocation, added occlusion, and changes to the task-relevant object set. These events require detecting stale beliefs and revising task progress accordingly. Each nominal instance is paired with six intervention-augmented variants derived from the same scene and goal, enabling controlled evaluation of state maintenance, exploration, and belief revision. More details are provided in the Appendix D.

Table 1: Symbolic open-loop tabletop results on Z-RoboSTAR-Bench. Each model is evaluated on 57 nominal scenes and 342 intervention scenes. SSAL is computed on nominal ones only.

	Nominal ($N = 57$)				Intervention ($N = 342$)		
Model	$\overline{N}_s \downarrow$	SR (%)↑	NGP (%)↑	SSAL↑	$\overline{N}_s \downarrow$	SR (%)↑	NGP (%)↑
Closed-source models							
DeepSeek-V4 Pro	46.6	54.4	87.0	0.406	50.8	48.2	85.6
GLM-5.2	53.3	47.4	78.8	0.324	55.7	43.0	77.7
Qwen3.6-Plus	39.7	28.1	78.4	0.224	40.8	19.9	74.7
Qwen3.7-Plus	43.1	29.8	77.2	0.232	47.7	23.4	72.8
GPT-5.4	65.7	26.3	69.3	0.190	70.2	22.5	67.4
Claude Opus 4.7	46.8	63.2	90.6	0.439	49.1	55.6	86.0
Gemini 3.1 Pro	40.5	64.9	87.3	0.529	**40.2**	62.3	87.2
GPT-5.5	**35.8**	**84.2**	**96.3**	**0.689**	40.7	**77.5**	**94.5**
Open-source models							
Qwen3-VL-8B	36.9	1.8	30.5	0.003	43.0	1.5	28.4
Qwen3-VL-32B	73.8	8.8	51.2	0.074	71.3	9.4	51.6
InternVL3.5-8B	96.8	1.8	34.4	0.003	95.9	2.3	31.0
InternVL3.5-38B	92.6	7.0	40.5	0.033	90.4	6.1	37.5
Cosmos-Reason1-7B	87.1	0.0	9.3	0.000	77.6	0.0	8.5
RoboBrain2.0-7B	57.1	0.0	4.8	0.000	58.5	0.0	3.4
RoboBrain2.0-32B	65.5	0.0	25.9	0.000	64.5	0.3	22.8
Our training model							
Z-RoboSTAR-27B	42.7	78.9	94.4	0.594	43.5	71.6	90.2

5 EVALUATION

5.1 EXPERIMENTAL SETUP

Closed-loop evluation. The proposed Z-RoboSTAR uses the same semantic action interface in all experiments, and only the executor that realizes an action and produces the next observation changes. We consider three execution modes: (i) *symbolic execution with open-loop observation*, in which Z-RoboSTAR-Bench's scene graph applies the semantic transition and returns the resulting symbolic state and collected visual observations in advance as an open-loop observation, (ii) *human execution*, in which an operator carries out the proposed action, and (iii) *VLA execution*, in which a VLA controller (e.g., a post-training $\pi_{0.5}$ policy in our evaluation) maps the semantic command directly to robot actions. The latter two modes practically connect to a Galaxea R1 Lite platform in the real world and use real-time camera observations from it. We train Z-RoboSTAR-27B using trajectories collected from symbolic open-loop execution. More details are provided in Appendix A.

Metrics. We report task success rate (SR), normalized goal progress (NGP), the mean number of non-stop actions $\overline{N}_s$, success weighted by symbolic action length (SSAL), and intervention recovery rate (IRR). NGP is the fraction of goal conditions satisfied at termination, while SSAL measures the efficiency of successful nominal rollouts relative to the shortest symbolic oracle. IRR is the fraction of rollouts that complete the task among those in which an intervention is actually applied, measuring recovery toward the original task goal. More details are provided in Appdendix E.

5.2 SYMBOLIC OPEN-LOOP AND HUMAN-EXECUTION CONSISTENCY

Main symbolic comparison. Table 1 compares 15 baseline models and Z-RoboSTAR-27B on 57 nominal and 342 intervention episodes per model. GPT-5.5 achieves the highest SR in both regimes, at 84.2% and 77.5%, with NGP above 94%. Z-RoboSTAR-27B ranks second, reaching 78.9% nominal SR and 71.6% intervention SR, with NGP of 94.4% and 90.2%, respectively. Among the open-source baselines, Qwen3-VL-32B has the highest SR (8.8% nominal and 9.4% under intervention). We find that open-source models, including those embodied foundation models, achieve low success rates on our sustained long-horizon tasks that need complex state tracking. Moreover, the dominant failure modes differ across models: closed-source models more often terminate prematurely after incorrectly judging the task to be complete, whereas open-source models tend to enter repetitive action loops that make no meaningful progress until they reach our predefined limit on the number of unproductive actions.

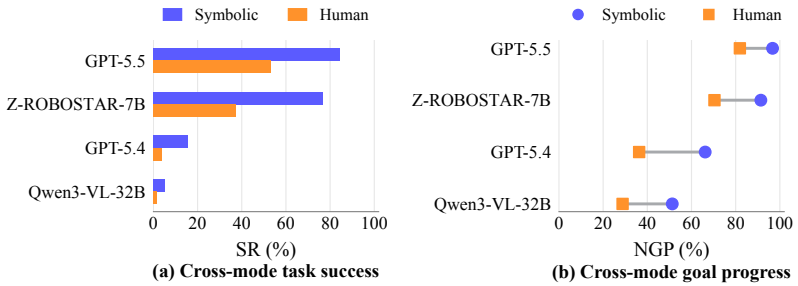


Figure 4: Cross-mode comparison on 57 tasks. Left: compare symbolic open-loop and human-executed closed-loop task success rates. Right: compare NGP under symbolic and human execution. Human sample counts are 79, 83, 78, and 71 for GPT-5.5, Z-RoboSTAR-27B, GPT-5.4, and Qwen3-VL-32B, respectively.

Spatial and temporal difficulty. We isolate intrinsic difficulty using nominal tasks only. Temporal difficulty is defined by the oracle action horizon. Spatial difficulty averages standardized graph node count, edge count, number of occluded nodes, and maximum occlusion depth. Empirical terciles define low, medium, and high groups. Table 2 shows a consistent decline in SR as temporal horizon increases, from 33.3% to 22.5%. High-spatial tasks are also harder than medium-spatial tasks (23.3% versus 32.2% SR), although the spatial trend is not monotonic.

Table 2: Nominal symbolic results by intrinsic difficulty, macro-averaged over 16 models.

Axis	Level	N	SR (%)	NGP (%)
Spatial	Low	33	28.5	56.6
	Medium	23	32.2	54.8
	High	28	23.3	50.7
Temporal	Low	29	33.3	56.6
	Medium	34	26.3	51.5
	High	21	22.5	50.6

Consistency under human execution. Figure 4 compares symbolic open-loop execution and human closed-loop execution as system 1 for four models across the same set of 57 tasks. Both SR and NGP preserve the same ordering: GPT-5.5, Z-RoboSTAR-27B, GPT-5.4, and Qwen3-VL-32B. Absolute performance drops under human execution, reflecting the additional visual and physical demands of acting in the real world.

Sensitivity to interventions. Table 3 reports performance by state change on the 57-scene tabletop split. Added occlusion causes the largest aggregate loss, reducing SR by 10.4 percentage points; adding an object and rolling back a completed subgoal reduce SR by 4.3 and 2.7 points, respectively. Wrong-container relocation yields a slightly higher pooled SR than the nominal condition, but this aggregate mixes triggered and untriggered episodes across heterogeneous model–task pairs and should not be interpreted as a benefit. IRR is therefore computed only over rollouts in which the intervention was actually applied.

Table 3: Symbolic open-loop results by intervention on 57 tabletop scenes, aggregated over 16 baseline models (i.e., 912 rollouts per condition). N_{app} counts triggered rollouts; ΔSR is measured in percentage points relative to the nominal aggregate; IRR scores recovery among triggered rollouts.

Condition	N_{app}	SR (%)	ΔSR	NGP (%)	IRR (%) ↑
Nominal	–	27.8	0.0	57.4	–
Action failure	840	26.0	−1.9	55.5	26.4
State regression	689	28.0	+0.1	57.0	33.7
Completed-subgoal rollback	744	25.1	−2.7	55.4	28.9
Unexpected relocation	731	28.8	+0.9	55.4	31.5
Added occlusion	518	17.4	−10.4	53.5	27.6
Added object	490	23.5	−4.3	54.8	41.0

5.3 Ablation Study

Experimental setting. We compare the original Z-RoboSTAR-27B baseline with three ablated variants: w/o global visual reconciliation, w/o task-progress monitoring, and w/o scene graph (Table 4). Each row aggregates partial human-execution trials in which a random human intervention is introduced at every execution step, continuously perturbing the evolving world state. Task SR

uses final goal verdicts, supplemented by retrospective adjudication when absent; NGP measures the fraction of goal conditions satisfied in the final logged state.

Table 4: Ablation comparisons of Z-RoboSTAR-27B on available human-execution trials. N counts retained trials; ΔSR is measured in percentage points relative to the baseline.

Variant	N	Task SR (%)	NGP (%)	ΔSR
Baseline Z-RoboSTAR-27B	83	37.3	78.8	0.0
w/o global visual reconciliation	23	21.7	73.9	-15.6
w/o task-progress monitoring	16	0.0	43.7	-37.3
w/o scene graph	23	4.3	38.4	-33.0

Results. On these trials, the baseline achieves 37.3% task SR and 78.8% NGP. Removing global visual reconciliation reduces performance to 21.7% task SR and 73.9% NGP. Without task-progress monitoring or the scene graph, task SR drops to 0.0% and 4.3%, corresponding to decreases of 37.3 and 33.0 percentage points, respectively. These ablations also lead to shorter execution horizons, further indicating that the agent struggles to sustain progress over extended interactions when persistent state maintenance is weakened.

5.4 End-to-End Robot Evaluation

We finally test whether state maintenance in System 2 improves physical execution when the System 1 controller is the same. The task is to organize an office desk, such as storing stationery, shelving books, discarding waste paper, etc. From the nominal initial configuration, a 23-action reference sequence satisfies all goals without interference or execution failures. Each system reported in Table 5 is evaluated over 11 trials. All systems use the same fine-tuned $\pi_{0.5}$ policy with our collected subtask trajectory data, such as `pick up the red pen`. VLA-only receives the complete task instruction, each Z-RoboSTAR+VLA condition issues semantic commands, and the oracle condition supplies the reference decomposition in advance that does not require a System 2 agent. This comparison holds the VLA constant while varying the source of the high-level planner.

Table 5: End-to-end robot results with a fixed fine-tuned $\pi_{0.5}$ controller. Action SR is measured by evaluator judgments and excludes `uncertain`. Task SR and time are computed over sessions with a recorded `session_end`. Parentheses report successful trials out of 11.

System	Action SR (%)	Task SR (%)	NGP (%)	Time (min)
VLA only	38.7	0 (0/11)	24.1	–
Z-RoboSTAR(GPT-5.5) + same VLA	43.1	81.8 (9/11)	95.5	64.8
Z-RoboSTAR(Z-RoboSTAR-27B) + same VLA	61.5	63.6 (7/11)	83.6	33.5
Oracle decomposition + same VLA	89.7	90.9 (10/11)	100.0	–

With the same System 1 policy, both embodied System 2 agent variants substantially outperform direct VLA execution at the task level: GPT-5.5 and Z-RoboSTAR-27B reach 81.8% and 63.6% task SR, respectively, whereas VLA-only reaches 0% for sustained long horizons. Their NGP scores similarly rise from 24.1% to 95.5% and 83.6%. Notably, GPT-5.5 attains higher task success despite a lower action SR than Z-RoboSTAR-27B, reinforcing that successful individual actions are not sufficient for sustained goal completion. The oracle result (90.9% task SR and 99.5% NGP) quantifies the remaining gap in high-level planning.

6 Conclusion

We presented Z-RoboSTAR, an embodied System 2 agent harness framework that uses a dynamically updated symbolic scene graph as a shared interface for scene belief, task progress, and subsequent planning. Across more than 6,384 symbolic rollouts on our Z-RoboSTAR-bench, GPT-5.5 achieved 84.2% nominal and 77.5% intervention success, while the matched human-execution study preserved the model ordering but exposed a substantial visual and physical execution gap. With the System 1 controller fixed, i.e., $\pi_{0.5}$, the GPT-5.5 and Z-RoboSTAR-27B brains reached 81.8% and 63.6% task success, compared with 0% for direct VLA execution, showing the value of maintaining and revising high-level state throughout our prolonged periods of execution. Overall, the results position synchronized scene-belief and task-progress revision as a central requirement for reliable and sustained long-horizon manipulation. We provide a demonstration in the supplementary material.

Ethics and Safety Statement

Our work studies long-horizon robotic manipulation using public and self-collected robotics data, human-executed evaluation, and real-robot experiments. Physical experiments are conducted under human supervision, and all released data will follow applicable privacy, consent, and licensing requirements.

Reproducibility Statement

We provide details of model training, agent interfaces, prompts, task construction, intervention settings, and evaluation protocols in the paper and appendix. Where permitted, we will release the corresponding code, configurations, checkpoints, and evaluation artifacts.

AI Use Statement

Generative AI tools were used to assist with manuscript organization and language editing.

References

Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691*, 2022.

Alisson Azzolini, Junjie Bai, Hannah Brandon, Jiaxin Cao, Prithvijit Chattopadhyay, Huayu Chen, Jinju Chu, Yin Cui, Jenna Diamond, Yifan Ding, et al. Cosmos-reason1: From physical common sense to embodied reasoning. *arXiv preprint arXiv:2503.15558*, 2025.

Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, et al. Gr00t n1: An open foundation model for generalist humanoid robots. *arXiv preprint arXiv:2503.14734*, 2025.

Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, and Chelsea Finn. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023.

Tianxing Chen, Yue Chen, Zixuan Li, Junyuan Tang, Kailun Su, Haoran Lu, Weijie Wan, Baijun Chen, Songling Liu, Haowen Yan, et al. Robodojo: A unified sim-and-real benchmark for comprehensive evaluation of generalist robot manipulation policies. *arXiv preprint arXiv:2607.04434*, 2026.

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.

Ronghao Dang, Jiayan Guo, Bohan Hou, Sicong Leng, Kehan Li, Xin Li, Jiangpin Liu, Yunxuan Mao, Zhikai Wang, Yuqian Yuan, et al. Rynnbrain: Open embodied foundation models. *arXiv preprint arXiv:2602.14979*, 2026.

Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*, 2023.

Jiafei Duan, Wilbert Pumacay, Nishanth Kumar, Yi Ru Wang, Shulin Tian, Wentao Yuan, Ranjay Krishna, Dieter Fox, Ajay Mandlekar, and Yijie Guo. Aha: A vision-language-model for detecting and reasoning over failures in robotic manipulation. *arXiv preprint arXiv:2410.00371*, 2024.

Yiguo Fan, Pengxiang Ding, Shuanghao Bai, Xinyang Tong, Yuyang Zhu, Hongchao Lu, Fengqi Dai, Wei Zhao, Yang Liu, Siteng Huang, et al. Long-vla: Unleashing long-horizon capability of vision language action model for robot manipulation. *arXiv preprint arXiv:2508.19958*, 2025.

Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. Inner monologue: Embodied reasoning through planning with language models. *arXiv preprint arXiv:2207.05608*, 2022.

Wenlong Huang, Chen Wang, Ruohan Zhang, Yunzhu Li, Jiajun Wu, and Li Fei-Fei. Voxposer: Composable 3d value maps for robotic manipulation with language models. *arXiv preprint arXiv:2307.05973*, 2023.

Wenlong Huang, Chen Wang, Yunzhu Li, Ruohan Zhang, and Li Fei-Fei. Rekep: Spatio-temporal reasoning of relational keypoint constraints for robotic manipulation. *arXiv preprint arXiv:2409.01652*, 2024.

Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, et al. $\pi_{0.5}$: a vision-language-action model with open-world generalization. *arXiv preprint arXiv:2504.16054*, 2025.

Yunfan Jiang, Agrim Gupta, Zichen Zhang, Guanzhi Wang, Yongqiang Dou, Yanjun Chen, Li Fei-Fei, Anima Anandkumar, Yuke Zhu, and Linxi Fan. Vima: Robot manipulation with multimodal prompts. *arXiv preprint arXiv:2302.03432*, 2023.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

Ruiying Li, Yunlang Zhou, YuYao Zhu, Kylin Chen, Jingyuan Wang, Sukai Wang, Kongtao Hu, Minhui Yu, Bowen Jiang, Zhan Su, et al. Roboclaw: An agentic framework for scalable long-horizon robotic tasks. *arXiv preprint arXiv:2603.11558*, 2026.

Yi Li, Yuquan Deng, Jesse Zhang, Joel Jang, Marius Memmel, Caelan Garrett, Fabio Ramos, Dieter Fox, Anqi Li, Abhishek Gupta, et al. Hamster: Hierarchical action models for open-world robot manipulation. In *International Conference on Learning Representations*, volume 2025, pp. 24040–24068, 2025.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. Libero: Benchmarking knowledge transfer for lifelong robot learning. *Advances in Neural Information Processing Systems*, 36:44776–44791, 2023.

Haowen Liu, Xirui Li, Shaoxiong Yao, Peng Shi, Tianyi Zhou, Jia-Bin Huang, Furong Huang, and Jiayuan Mao. Guava: An effective and universal harness for embodied manipulation. *arXiv preprint arXiv:2606.18363*, 2026a.

Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *European conference on computer vision*, pp. 38–55. Springer, 2024.

Yicheng Liu, Zibin Dong, Baijun Ye, Tianyuan Yuan, Tao Jiang, Anqi Yang, Shicheng Cao, Haonan Liu, Yue Sun, Zihan Guo, et al. G0. 5: One autoregressive stream for robot reasoning and action. *arXiv preprint arXiv:2608.11739*, 2026b.

Yuanzhe Liu, Jingyuan Zhu, Yuchen Mo, Gen Li, Xu Cao, Jin Jin, Yifan Shen, Zhengyuan Li, Tianjiao Yu, Wenzhen Yuan, et al. Palm: Progress-aware policy learning via affordance reasoning for long-horizon robotic manipulation. *arXiv preprint arXiv:2601.07060*, 2026c.

Oier Mees, Lukas Hermann, Erick Rosete-Beas, and Wolfram Burgard. Calvin: A benchmark for language-conditioned policy learning for long-horizon robot manipulation tasks. *IEEE Robotics and Automation Letters*, 7(3):7327–7334, 2022.

Soroush Nasiriany, Abhiram Maddukuri, Lance Zhang, Adeet Parikh, Aaron Lo, Abhishek Joshi, Ajay Mandlekar, and Yuke Zhu. Robocasa: Large-scale simulation of everyday tasks for generalist robots. *arXiv preprint arXiv:2406.02523*, 2024.

Maëlic Neau. Relateanything: Real-time open-vocabulary relation prediction from any inputs. *arXiv preprint arXiv:2609.12552*, 2026.

OpenAI. GPT-6 Astra. https://developers.openai.com/api/docs/models/gpt-6-astra, 2026. Accessed: 2026-09-22.

Qwen Team. Qwen3-VL: Sharper Vision, Deeper Thought, Broader Action. https://qwen.ai/blog?id=99f0335c4ad9ff6153e517418d48535ab6d8afef, 2025. Accessed: 2026-09-22.

Krishan Rana, Jesse Haviland, Sourav Garg, Jad Abou-Chakra, Ian Reid, and Niko Suenderhauf. Sayplan: Grounding large language models using 3d scene graphs for scalable robot task planning. *arXiv preprint arXiv:2307.06135*, 2023.

Lucy Xiaoyang Shi, Brian Ichter, Michael Equi, Liyiming Ke, Karl Pertsch, Quan Vuong, James Tanner, Anna Walling, Haohuan Wang, Niccolo Fusai, et al. Hi robot: Open-ended instruction following with hierarchical vision-language-action models. *arXiv preprint arXiv:2502.19417*, 2025.

Huajie Tan, Enshen Zhou, Zhiyu Li, Yijie Xu, Yuheng Ji, Xiansheng Chen, Cheng Chi, Pengwei Wang, Huizhu Jia, Yulong Ao, et al. Robobrain 2.5: Depth in sight, time in mind. *arXiv preprint arXiv:2601.14352*, 2026.

Gemini Robotics Team, Abbas Abdolmaleki, Saminda Abeyruwan, Joshua Ainslie, Jean-Baptiste Alayrac, Montserrat Gonzalez Arenas, Ashwin Balakrishna, Nathan Batchelor, Alex Bewley, Jeff Bingham, et al. Gemini robotics 1.5: Pushing the frontier of generalist robots with advanced embodied reasoning, thinking, and motion transfer. *arXiv preprint arXiv:2510.03342*, 2025.

Qi Wang, Tianyi Wang, Chengyang Li, Shikun Ban, Yurun Chen, Yizhong Ge, Jason Qin, Chengtai Li, and Wentao Zhu. Towards the harness of embodied agents. *arXiv preprint arXiv:2608.11246*, 2026.

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*, 2025.

Borong Zhang, Jiahao Li, Jiachen Shen, Yuhao Zhang, Yishuai Cai, Lu Liu, Hailu Ji, Yuanpei Chen, Juntao Dai, Jiaming Ji, et al. Vla-arena: An open-source framework for benchmarking vision-language-action models. *arXiv preprint arXiv:2512.22539*, 2025a.

Shiduo Zhang, Zhe Xu, Peiju Liu, Xiaopeng Yu, Yuan Li, Qinghui Gao, Zhaoye Fei, Zhangyue Yin, Zuxuan Wu, Yu-Gang Jiang, et al. Vlabench: A large-scale benchmark for language-conditioned robotics manipulation with long-horizon reasoning tasks. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11142–11152. IEEE, 2025b.

Xiaolin Zhou, Liu Liu, Tingyang Xiao, Wei Feng, Fa Fu, Xinrui Meng, Xinjie Wang, Jialiang Han, Boyang Yu, Yun Du, et al. Holoagent-0: A unified embodied agent framework with 3d spatial memory. *arXiv preprint arXiv:2606.23565*, 2026.

Zihan Zhou, Animesh Garg, Dieter Fox, Caelan Garrett, and Ajay Mandlekar. Spire: Synergistic planning, imitation, and reinforcement learning for long-horizon manipulation. *arXiv preprint arXiv:2410.18065*, 2024.

A Evaluation Platform and Execution Protocol

Shared System 2 to System 1 interface. All three execution modes consume the same semantic command representation: an action name and grounded entity arguments. The execution mode determines how the action is realized and how the next observation is obtained. Table 6 summarizes this boundary.

Table 6: Execution modes used in the evaluation.

Mode	Action realization	Observation and feedback
Symbolic open-loop execution	Executable action schemas update the scene graph state.	Symbolic state and previously collected visual observations provide the configured observation stream. The System 2 brain evaluates, reconciles, and continues or replans.
Human execution	An operator carries out the proposed semantic action.	The robot returns new camera observations after an execution window. The System 2 brain evaluates, reconciles, and continues or replans.
System 2 brain + VLA (System 1)	A fine-tuned $\pi_{0.5}$ policy executes the command through its registered skill prompt.	The robot returns new camera observations after an execution window. The System 2 brain evaluates, reconciles, and continues or replans.

Physical observations. Human and VLA execution use the Galaxea R1 Lite platform with a head camera and two wrist cameras. Before/after visual evidence used in the evaluator and reconciliation is compared within matched camera pairs. The complete example in Appendix F retains all three views at every observation history boundary, including unchanged views and failed actions.

Semantic and motor time scales. A semantic subtask command can span several low-level action chunks. The policy's action horizon specifies the number of predictions in one chunk, whereas the execution-window size specifies how many chunks are consumed before the next visual judgment. These quantities are recorded separately. An `in_progress` judgment permits another execution window for the same pending action; `uncertain` requests further observation without treating the action as completed. A failure returns control to recovery or replanning.

Termination and retained human trials. The harness records explicit stop or termination events and the reason for termination. In the human-execution aggregates, retained trials/sessions have a recorded terminal event or more than 30 completed success/failure judgments. Long nonterminated trials are counted as failures. Short interrupted starts are excluded. Uncertain re-observations do not increase the completed-step count. This session-selection rule is separate from the within-session step budget and stagnation checks. Per-step outcomes, task verdicts, and graph commits are recorded separately.

B VLA Subtasks and Training

B.1 Deployed System 1 policy

The System 1 controller leveraging the VLA policy streams observations and executes joint commands. The policy is a $\pi_{0.5}$ model trained in two successive stages: OpenGalaxea training from the released $\pi_{0.5}$ base for 100k steps, followed by joint training on our desk-manipulation subtasks for 50k steps.

Data and sampling. The OpenGalaxea corpus uses the R1 Lite embodiment and English instructions. Its merge manifest contains 226 retained source repositories after robot-type, timestamp, and invalid-video filtering, with 20,623 episodes and 26,254,215 frames at 15 Hz. Stage 2 uses only the 23 curated task (i.e., Desk23) repositories listed in Table 8, sampled with probability $1/23$ per

Table 7: Training configuration of the deployed System 1 controller. Batch sizes are global.

Setting	Stage 1: OpenGalaxea	Stage 2: Desk23
Initialization	Released $\pi_{0.5}$ base	Stage-1 checkpoint 99999
Training updates	100,000	50,000
Final checkpoint index	99999	49999
Global batch size	768	256
Hardware allocation	12 nodes, 96 GPUs	1 node, 8 GPUs
Batch per GPU	8	32
FSDP group size	8 devices	8 devices
Learning-rate warmup	2,000 updates	1,000 updates
Peak learning rate	4×10^{-5}	2.5×10^{-5}
Cosine decay horizon	100,000 updates	50,000 updates
Final learning rate	4×10^{-6}	2.5×10^{-6}
Demonstration episodes	20,623	2,372
Recorded frames	26,254,215	428,035
Data repositories	226 source repositories	23 task repositories
Distinct instruction strings	3,914	22
Sampling mixture	Merged OpenGalaxea corpus	Equal weight per task repository
Checkpoint interval	5,000 updates	5,000 updates

repository. No OpenGalaxea replay mixture is configured in this second stage. The training budgets correspond to 76.8 million and 12.8 million sampled examples, respectively, including repeated samples.

Optimization and policy representation. Both stages use full-parameter $\pi_{0.5}$ training with the `gemma_2b` vision-language backbone and `gemma_300m` action expert, bfloat16 model computation, AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.95$, $\epsilon = 10^{-8}$, weight decay 10^{-10}), global gradient-norm clipping at 1.0, and an exponential moving average with decay 0.99. The random seed is 42. Each prediction contains 16 action steps with a padded action dimension of 32. The robot interface uses 14 coordinates: six joints and one gripper coordinate per arm. Arm targets are trained as offsets from the current joint state and converted back to absolute joint targets for execution; gripper commands remain absolute and are scaled to $[0, 1]$ in the policy interface (Intelligence et al., 2025).

Visual inputs and normalization. The policy receives the current head, left-wrist, and right-wrist images, the robot state, and the subtask instruction. Images are resized with padding to 224×224. Both stages use the `strong` training augmentation: the head view receives a 0.95-scale random crop and up to 5° rotation; all views receive color jitter with brightness, contrast, and saturation magnitudes of 0.3, 0.4, and 0.5. Augmentation is disabled during inference. State/action normalization uses the first and 99th percentiles. Stage 2 recomputes these statistics with equal task-repository weights and deploys the Desk23 statistics with its checkpoint.

B.2 Complete Desk23 subtask inventory

Runtime action bindings. The 23 curated task (i.e., Desk23) repositories are listed in Table 8. The strict prompt catalog of action commands grounds these instructions to exact symbolic action/entity pairs. For example, pick-up instructions implement `grab`, placements implement `putin`, opening, closing, and attachment implement `open`, `close`, and `attach`. In total, the recorded catalog has 29 grounded bindings and 25 distinct instruction strings.

C Asynchronous System Execution

System 1 and System 2. System 1 collects the head and wrist-camera observations and the robot state, and executes the VLA action chunks at 15 Hz. System 2 selects semantic actions, evaluates their outcomes, reconciles the scene belief, and updates task progress. Observation exchange and inference run asynchronously from the System 1 control loop; while waiting for a response, System 1 holds the robot pose.

Table 8: All 23 task repositories used to train the deployed System 1 controller. Rows 20 and 21 are separate collections with the same instruction. Frames are measured at 15 Hz.

ID	VLA subtask instruction	Episodes	Frames
1	pick up the red pen	102	13,949
2	pick up the black pen	105	13,063
3	pick up the blue pen	103	13,771
4	put the red pen into the pencil case	105	10,896
5	put the black pen into the pencil case	103	14,844
6	put the blue pen into the pencil case	102	13,356
7	pick up the eraser	102	17,858
8	open the top drawer	102	23,115
9	close the top drawer	104	16,354
10	open the second drawer	103	18,639
11	close the second drawer	103	17,064
12	put the eraser into the pencil case	103	13,837
13	open the third drawer	103	18,513
14	close the third drawer	105	16,826
15	attach the cap on the lipstick	102	51,627
16	pick up the waste paper	103	13,598
17	put the waste paper into the desk trash bin	103	17,069
18	close the pencil case	102	23,303
19	open the pencil case	102	21,259
20	open the book	102	24,464
21	open the book	102	22,315
22	pick up the book	106	18,818
23	put the book into the first shelf compartment	105	13,497
	Total	2,372	428,035

Synchronization and feedback. Each VLA chunk contains 16 control steps. In the recorded configuration, System 1 consumes the current chunk and waits for motion to settle before requesting the next chunk. The default semantic execution window is eight chunks, with four chunks for opening either book. Within a window, System 1 returns fresh observations while retaining the same semantic instruction. At the window boundary, System 2 evaluates the action and reconciles the scene belief before continuing the action or selecting the next one. An `in_progress` result can extend execution of the pending action; an `uncertain` result requests another observation. System 1 ends execution when System 2 returns a terminal response.

D Task Compilation and Intervention Suite of Z-RoboSTAR-Bench

Tabletop tasks. The 57-scene tabletop benchmark contains six task families (Table 9). Each scene is paired with one nominal condition and six intervention conditions, giving 399 assigned rollouts per model.

Table 9: Scene counts in the 57-scene tabletop benchmark.

Task family	Scenes
Desk organization	10
Cosmetics organization	10
Toy organization	10
Dining-table organization	10
Loading condiments into a refrigerator	10
Dividing buffet trays	7
Total	57

Compiled instance. Following Section 4, a compiled instance is

$$\mathcal{I} = (G_0^\star, g, \mathcal{A}, \mathcal{O}, \Lambda),$$

where $G_0^\star$ is the initial benchmark scene graph, g specifies the goal predicates, $\mathcal{A}$ defines semantic action schemas, $\mathcal{O}$ governs partial observations, and Λ is the set of exogenous events.

Spatial and temporal dependencies. Spatial dependencies place task-relevant facts outside the current observation, such as inside closed containers or behind occluders, requiring active exploration before they can inform later decisions. Temporal dependencies arise when later steps depend on the completion or information acquired from earlier steps, creating causal chains across the task.

Dynamic interventions. As in Section 4, an intervention is $\lambda = \langle \psi_\lambda, \delta_\lambda, \kappa_\lambda \rangle$, comprising a semantic trigger, a graph rewrite, and an execution condition. Triggers depend on task progress rather than wall-clock time. Each nominal instance is paired with six intervention variants derived from the same scene and goal.

Table 10: The six assigned intervention families. An intervention is applied only when its semantic trigger and execution condition are met.

Intervention	Change the agent must detect or recover from
Action failure	The requested action does not produce its intended state transition.
State regression	A previously advanced state is reverted, invalidating the corresponding belief or prerequisite.
Completed-subgoal rollback	A satisfied goal condition becomes unsatisfied and must be pursued again.
Unexpected relocation	An object is moved to a different container, requiring its location and subsequent plan to be revised.
Added occlusion	A new obstruction changes which task-relevant entities are observable or accessible.
Added object	The task-relevant object set changes, introducing additional state or goal obligations.

Assigned versus applied interventions. An assigned intervention condition does not imply that its trigger fires in every rollout. Condition-level SR and NGP include all rollouts assigned to that condition. IRR instead conditions on an intervention actually being applied.

E Metric Definitions and Aggregation

This section defines the five metrics reported in Section 5: task success rate (SR), normalized goal progress (NGP), the mean number of non-stop actions $\overline{N}_s$, success weighted by symbolic action length (SSAL), and intervention recovery rate (IRR).

Task success and goal progress. Let N be the number of evaluated rollouts and $y_i \in \{0, 1\}$ indicate whether rollout i completes the task. Let K_i be its number of goal conditions at termination and $b_{ij} \in \{0, 1\}$ indicate whether condition j is satisfied in the final state. Then

$$\mathrm{SR} = \frac{100}{N} \sum_{i=1}^{N} y_i, \qquad \mathrm{NGP} = \frac{100}{N} \sum_{i=1}^{N} \frac{\sum_{j=1}^{K_i} b_{ij}}{K_i}.$$

NGP averages the fraction of satisfied goal conditions across rollouts, including partially completed tasks.

Action count and nominal efficiency. Let n_i be the number of non-stop actions in rollout i and $\ell_i^\star$ the shortest symbolic oracle action length for its nominal task. We report

$$\overline{N}_s = \frac{1}{N} \sum_{i=1}^{N} n_i, \qquad \mathrm{SSAL} = \frac{1}{N} \sum_{i=1}^{N} y_i \frac{\ell_i^\star}{\max(\ell_i^\star, n_i)}.$$

SSAL is computed on nominal rollouts only. Unsuccessful rollouts contribute zero; successful rollouts are weighted by their action length relative to the oracle.

Intervention recovery. Let $a_i \in \{0, 1\}$ indicate whether an intervention was actually applied in rollout i. Then

$$\mathrm{IRR} = 100 \frac{\sum_i a_i y_i}{\sum_i a_i}.$$

IRR measures task completion among rollouts with an applied intervention. It is undefined when no intervention is applied. Assigned intervention conditions and actually applied interventions therefore use different denominators.

F Z-RoboSTAR AND VLA EXECUTION EXAMPLE

The following pages reproduce 15 fully recorded actions from a System 2 execution of Z-RoboSTAR, with the VLA policy as System 1 for desk organization. In the initial observation, the black pen lies on the desktop beside the open pencil case. Every displayed action includes an action plan, an evaluator response, and a global visual reconciliation response.

Table 11: Configuration recorded for Z-RoboSTAR and VLA session. The policy is described in Appendix B; synchronization is described in Appendix C.

Field	Recorded value
VLA action horizon	16 control steps per chunk
Execution window	Eight chunks; four for opening the book
Reconciliation mode	Enabled
Reconciliation confidence threshold	0.8
Reconciliation response format	Strict JSON schema
Action-validation attempts	3
Stagnation limit	10 completed steps without improving best cost

Each action is paired with the head, left-wrist, and right-wrist observations before and after execution, its complete action-plan and evaluator JSON, selected backend commit fields, and the complete global visual reconciliation response. JSON structure, numbers, Booleans, and model verdicts are retained. These are logged model outputs, including errors and disagreements between the evaluator and reconciliation.

F.1 STEP 1

Action: `grab(black_pen)`

Logged action ID: `58fe1f1d538b47f1bde46ceb829fc4bf`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The pencil case is open, and the black pen is on the desktop and needs to be
  placed inside the pencil case. Pick up the black pen first.",
  "action": {"name": "grab", "object": "black_pen", "target": null, "node_ids": ["black_pen"]}
}
```


Evaluator JSON.


```
{"status": "failure", "reason": "The black pen was moved/placed into the purple tray, but it
  is not visibly grasped by the gripper after the action; the target-specific result of
  grabbing the black pen is not achieved."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": false,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 19.0, "best_cost": 19.0, "stagnant_steps": 0, "
  max_stagnant_steps": 10, "last_cost_change_step": 1, "enabled": true, "cost_before": 21.0, "
  cost_after": 19.0, "cost_changed": true, "improved_historical_best": true},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```


Step 1: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
   AFTER shows the white drawer cabinet at left; counting fixed drawer fronts from the cabinet
   top, the top slot is visible above the second and third fronts.", "state_evidence": "
  cam_high AFTER shows the top drawer front seated flush with the cabinet face, with no tray
  protruding."},
```

```
    "drawer_2": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
     AFTER shows the same cabinet; the second slot is the middle drawer front between the top
    and bottom fronts.", "state_evidence": "cam_high AFTER shows the middle drawer front seated
    flush, with its handle aligned and no pulled-out tray."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
     AFTER shows the same cabinet; the third slot is the bottom drawer front below the middle
    front.", "state_evidence": "cam_high AFTER shows the bottom drawer front seated flush, with
    no visible displacement from the cabinet."}
   },
   "object_observations": {
    "book_1": {"status": "consistent", "evidence": "cam_high AFTER shows one closed book on
    the desktop near the front-right area; books are interchangeable, assigned to book_1,
    matching ON desktop and closed."},
    "book_2": {"status": "consistent", "evidence": "cam_high AFTER and cam_right_wrist AFTER
    show a second closed book resting on the desktop to the right/front; books are
    interchangeable, assigned to book_2, matching ON desktop and closed."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue shelf/
    rack standing on the desktop in the expected area; no contradictory state or placement is
    observed."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
     shelf compartments on the desktop; the first compartment remains part of the shelf
    structure with no observed mismatch."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
     shelf compartments on the desktop; the third compartment remains part of the shelf
    structure with no observed mismatch."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
     shelf compartments on the desktop; the second compartment remains part of the shelf
    structure with no observed mismatch."},
    "lipstick": {"status": "unobservable", "evidence": "The assembled lipstick object as a
    whole is not directly identifiable in AFTER views; only a lipstick-like tube/body is visible
    , so no full-object location can be established."},
    "lipstick_cap": {"status": "unobservable", "evidence": "No distinct lipstick cap is
    directly visible in the AFTER views; the believed hidden/no-location state cannot be
    visually checked."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high AFTER shows the lipstick
    tube/body upright on the desktop near the eraser, matching belief ON desktop."},
    "waste_paper": {"status": "unobservable", "evidence": "No waste paper is directly visible
    in AFTER views; if under/inside a book or otherwise occluded, its current location cannot be
     established."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
    drawer cabinet in place on the desktop; no relocation or overall cabinet state mismatch is
    observed."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
     layer and shows it closed flush, matching current belief open=false."},
    "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
    drawer layer and shows it closed flush, matching current belief open=false."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
    drawer layer and shows it closed flush, matching current belief open=false."},
    "desktop": {"status": "consistent", "evidence": "All AFTER views show the white desktop
    surface supporting the visible objects; no mismatch for the static surface is observed."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the cream
    trash bin with black rim resting on the desktop at the back-right, matching ON desktop."},
    "eraser": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist AFTER
    show the small beige eraser resting on the desktop near the pencil box, matching ON desktop
    ."},
    "red_pen": {"status": "unobservable", "evidence": "No red pen is directly visible in AFTER
     views; the relevant closed middle drawer contents are not visible, so hidden/no-location
    belief cannot be checked."},
    "blue_pen": {"status": "unobservable", "evidence": "No blue pen is directly visible in
    AFTER views; the relevant closed bottom drawer contents are not visible, so hidden/no-
    location belief cannot be checked."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
    AFTER show the purple pencil box open on the desktop, matching belief ON desktop and open=
    true."},
    "black_pen": {"status": "correction_needed", "evidence": "cam_left_wrist AFTER clearly
    shows the black pen lying inside the open purple pencil box tray, not directly ON the
    desktop as current belief states."}
   },
   "reason": "AFTER images show the attempted grab failed: the black pen remains visible but is
     now clearly inside the open pencil box rather than on the bare desktop. The drawer cabinet
    is fully visible in cam_high AFTER and all three numbered layers are closed flush. Other
    visible catalog objects agree with their believed placements; hidden items are not changed
    without direct evidence.",
   "observed_state_delta": [
    {"operation": "relocate", "node_id": "black_pen", "relation": "INSIDE", "target": "
    pencil_case", "confidence": 0.95, "evidence": "cam_left_wrist AFTER shows the identified
    black pen resting within the purple pencil box tray; cam_high AFTER also shows the pen at
    the pencil box area, not on the bare desktop."}
   ]
}
```

F.2 STEP 2

Action: `grab(eraser)`

Logged action ID: `8f23a9cc9dc248b1a148ef69ef567f12`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The pencil case is open, and the eraser is visible on the desktop and needs to be
    placed inside the pencil case. Pick up the eraser first.",
  "action": {"name": "grab", "object": "eraser", "node_ids": ["eraser"]}
}
```


Evaluator JSON.


```
{"status": "failure", "reason": "The eraser remains on the table in the after images and is
   not visibly grasped by the gripper."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": false,
  "observation_patch_committed": false,
  "graph_update_committed": false,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 19.0, "best_cost": 19.0, "stagnant_steps": 1, "
  max_stagnant_steps": 10, "last_cost_change_step": 1, "enabled": true, "cost_before": 19.0, "
  cost_after": 19.0, "cost_changed": false, "improved_historical_best": false},
  "termination_triggered": false,
  "reconciliation_status": "ignored",
  "validation_errors": []
}
```


Step 2: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.93, "identity_evidence": "
  cam_high AFTER shows the white drawer cabinet at the left; counting fixed fronts from the
  cabinet top, the top drawer slot is layer 1 above the other two fronts.", "state_evidence":
  "cam_high AFTER shows the top drawer front seated flush with the cabinet, with no tray
  pulled out."},
```

```
    "drawer_2": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
   cam_high AFTER shows the same cabinet; the middle front between the top and bottom drawer
   fronts is layer 2.", "state_evidence": "cam_high AFTER shows the middle drawer front flush
   in the frame, not displaced open."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.91, "identity_evidence": "
   cam_high AFTER shows the same cabinet; the lowest visible front below the middle slot is
   layer 3.", "state_evidence": "cam_high AFTER shows the bottom drawer front seated flush,
   with no exposed tray."}
   },
   "object_observations": {
    "book_1": {"status": "consistent", "evidence": "cam_high AFTER and cam_right_wrist AFTER
   show two books on the desktop; assigning one interchangeable book to the left/front book, it
    remains ON desktop and appears closed."},
    "book_2": {"status": "consistent", "evidence": "cam_high AFTER and cam_right_wrist AFTER
   show the second interchangeable book on the desktop at the right/front area, also resting ON
    desktop and closed."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue book rack
    stationary on the desktop; its furniture placement is not contradicted."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    rack compartments on the desktop; the first compartment is visible as part of the rack and
   no contradictory movement is observed."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    rack compartments on the desktop; the third compartment remains part of the rack with no
   observed mismatch."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    rack compartments on the desktop; the second compartment remains part of the rack with no
   observed mismatch."},
    "lipstick": {"status": "unobservable", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show the lipstick tube/body area on the desktop, but the assembled whole lipstick with
    null location is not separately established."},
    "lipstick_cap": {"status": "unobservable", "evidence": "No AFTER view clearly shows a
   separate lipstick cap or the inside of the top drawer; its believed hidden/unknown location
   cannot be checked."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show the lipstick body standing on the desktop near the eraser, matching ON desktop
   ."},
    "waste_paper": {"status": "unobservable", "evidence": "The paper believed occluded by a
   book is not directly visible; the relevant book/under-book area is not opened or clearly
   inspected in AFTER views."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   drawer cabinet stationary on the tabletop; no furniture relocation or state mismatch for the
    cabinet itself is observed."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
    front as layer 1 and shows it flush closed, matching the believed closed state."},
    "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the lowest
   drawer front as layer 3 and shows it flush closed, matching the believed closed state."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
   drawer front as layer 2 and shows it flush closed, matching the believed closed state."},
    "desktop": {"status": "consistent", "evidence": "All AFTER views show the white desktop
   surface supporting the visible objects; no contradiction to its static surface belief."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER and
   cam_right_wrist AFTER show the white desktop trash bin resting on the desktop at the back
   right, matching ON desktop."},
    "eraser": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist AFTER
   show the small beige eraser resting on the desktop between the pencil box and lipstick; it
   is not held, matching ON desktop after the failed grab."},
    "red_pen": {"status": "unobservable", "evidence": "No AFTER view opens or clearly shows
   the second drawer interior; the red pen's hidden/unknown placement cannot be verified."},
    "blue_pen": {"status": "unobservable", "evidence": "No AFTER view opens or clearly shows
   the third drawer interior; the blue pen's hidden/unknown placement cannot be verified."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show the purple pencil case on the desktop with its lid open, matching ON desktop and
   open=true."},
    "black_pen": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist AFTER
    show the black pen lying inside the open purple pencil case, matching INSIDE pencil_case."}
   },
   "reason": "AFTER images show the attempted grab did not move the eraser: it remains on the
   desktop and the grippers are empty. The high camera clearly establishes all three drawer
   layers as closed and flush. Other visible objects match their current believed supports;
   hidden drawer contents and occluded paper are not directly verifiable.",
   "observed_state_delta": []
}
```

F.3 STEP 3

Action: `grab(book_1)`

Logged action ID: `069c240512e34301b098669f3b99b5fa`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "book_1 is visible on the desktop and needs to be placed in shelf_compartment_1.
  Both hands are currently free, so pick up book_1 first.",
  "action": {"name": "grab", "object": "book_1", "node_ids": ["book_1"]}
}
```


Evaluator JSON.


```
{"status": "success", "reason": "The target book on the right is visibly lifted from its
  original position and held between the gripper jaws in the after images."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": true,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 16.0, "best_cost": 16.0, "stagnant_steps": 0, "
  max_stagnant_steps": 10, "last_cost_change_step": 3, "enabled": true, "cost_before": 19.0, "
  cost_after": 16.0, "cost_changed": true, "improved_historical_best": true},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```


Step 3: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
   AFTER shows the white drawer cabinet on the left; the top drawer slot directly under the
  open top bin is distinguishable from the two lower fronts.", "state_evidence": "cam_high
  AFTER shows the first-layer front seated flush with the cabinet, with no tray pulled out."},
    "drawer_2": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
   AFTER shows the same cabinet; the second drawer is the middle front between the top and
```

```
  bottom drawer fronts.", "state_evidence": "cam_high AFTER shows the second-layer front
  seated flush, not displaced outward."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
   AFTER shows the same cabinet; the third drawer is the bottom front below the middle front
  .", "state_evidence": "cam_high AFTER shows the third-layer front seated flush, not pulled
  out."}
  },
  "object_observations": {
    "book_1": {"status": "consistent", "evidence": "cam_high and cam_right_wrist AFTER show
  one book held upright in the right gripper at the right side; books are interchangeable, so
  this matches the believed held book slot."},
    "book_2": {"status": "consistent", "evidence": "cam_high and cam_right_wrist AFTER show
  the other book still resting flat on the tabletop near the center-right, matching ON desktop
  ."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue book rack
   stationary on the desktop; no relocation is indicated for this furniture."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
   rack compartments on the tabletop; the first slot remains part of the rack on the desk,
  with no contradicting placement evidence."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
   rack compartments on the tabletop; the third slot remains part of the rack on the desk,
  with no contradicting placement evidence."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
   rack compartments on the tabletop; the second slot remains part of the rack on the desk,
  with no contradicting placement evidence."},
    "lipstick": {"status": "unobservable", "evidence": "No assembled lipstick object is
  distinctly visible in AFTER views; only the tube body is visible on the desk, so the catalog
   object lipstick cannot be localized."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not show a distinct
   lipstick cap; drawer interiors are not visible, so its current location cannot be checked
  ."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
  show the lipstick tube body standing on the tabletop near the eraser, matching ON desktop
  ."},
    "waste_paper": {"status": "correction_needed", "evidence": "cam_right_wrist AFTER shows
  crumpled waste paper exposed on the tabletop behind/right of the lifted book, previously
  hidden by that book; it is visibly resting ON desktop."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
  drawer cabinet stationary on the tabletop; no furniture relocation is observed."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
   layer and shows its front flush closed, consistent with the believed closed state."},
    "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
  drawer layer and shows its front flush closed, consistent with the believed closed state."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
  drawer layer and shows its front flush closed, consistent with the believed closed state."},
    "desktop": {"status": "consistent", "evidence": "All AFTER cameras show the white tabletop
   as the supporting surface for the stationary items; no mismatch for the surface itself."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the white/
  black-rim trash bin on the tabletop at the back right, matching ON desktop."},
    "eraser": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER show the
   small beige eraser resting on the tabletop near the lipstick tube, matching ON desktop."},
    "red_pen": {"status": "unobservable", "evidence": "AFTER views do not show a red pen;
  drawer interiors are closed/not visible, so its hidden location cannot be checked."},
    "blue_pen": {"status": "unobservable", "evidence": "AFTER views do not show a blue pen;
  drawer interiors are closed/not visible, so its hidden location cannot be checked."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
  show the purple pencil box open on the tabletop, consistent with believed ON desktop and
  open state."},
    "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows a black/
  white pen lying inside the open purple pencil box, matching INSIDE pencil_case."}
  },
  "reason": "AFTER images show the grab succeeded: one interchangeable book is held in the
  gripper while the other remains on the desk. Lifting the right-side book exposes crumpled
  waste paper on the tabletop. The drawer cabinet is fully visible in cam_high AFTER and all
  three numbered drawer fronts are flush closed. Other visible desk objects match their
  believed placements; hidden drawer contents remain unobservable.",
  "observed_state_delta": [
    {"operation": "relocate", "node_id": "waste_paper", "relation": "ON", "target": "desktop",
   "confidence": 0.86, "evidence": "cam_right_wrist AFTER shows the area previously covered by
   the lifted book; a crumpled paper is clearly visible resting on the white tabletop to the
  right/back of the remaining flat book."},
    {"operation": "reveal", "source": "book_1", "node_id": "waste_paper", "confidence": 0.86,
  "evidence": "cam_right_wrist AFTER shows the lifted/held book no longer covering its prior
  tabletop area, revealing the crumpled waste paper behind/right of it."}
  ]
}
```

F.4 STEP 4

Action: `putin(book_1,shelf_compartment_1)`

Logged action ID: `59fb2ed4362a4a2bba386e9736c9747d`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "book_1 is currently held, and shelf_compartment_1 is visible. The goal requires
  book_1 to be placed in shelf_compartment_1, so complete this placement first.",
  "action": {"name": "putin", "object": "book_1", "target": "shelf_compartment_1", "node_ids":
   ["book_1", "shelf_compartment_1"]}
}
```


Evaluator JSON.


```
{"status": "failure", "reason": "The book is not placed into the blue bookshelf first slot;
  after images show the target book removed from view/area and the gripper holding a crumpled
  paper instead, with no visible successful insertion into the shelf."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": false,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 16.0, "best_cost": 16.0, "stagnant_steps": 1, "
  max_stagnant_steps": 10, "last_cost_change_step": 3, "enabled": true, "cost_before": 16.0, "
  cost_after": 16.0, "cost_changed": false, "improved_historical_best": false},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```


Step 4: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.93, "identity_evidence": "
  cam_high AFTER shows the white three-slot drawer cabinet at left; the top drawer front is
  the first slot below the cabinet rim, above the two lower handles.", "state_evidence": "
```

```
cam_high AFTER shows the top drawer front seated flush with the cabinet, with no tray
protruding."},
  "drawer_2": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
cam_high AFTER shows the same cabinet; the second drawer is the middle front between the top
 and bottom drawer handles.", "state_evidence": "cam_high AFTER shows the middle drawer
front flush and not pulled out."},
  "drawer_3": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
cam_high AFTER shows the same cabinet; the third drawer is the bottom front below the middle
 handle, above the tabletop.", "state_evidence": "cam_high AFTER shows the bottom drawer
front flush and no exposed tray."}
},
"object_observations": {
  "book_1": {"status": "correction_needed", "evidence": "Books are interchangeable. cam_high
 AFTER shows the previously gripped upright book no longer in the visible gripper area and
instead absent from the gripper; cam_right_wrist AFTER shows the gripper holding crumpled
paper, not a book. No clear final book placement is visible, so held=true is contradicted
but no relocate is supported."},
  "book_2": {"status": "consistent", "evidence": "Using interchangeable book assignment,
cam_high AFTER shows one book lying on the tabletop near the front-center/right, matching
belief that a book is ON desktop; its exact original ID is not required."},
  "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue vertical
file/book shelf standing on the tabletop; static furniture placement is visually consistent
where checkable."},
  "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the left
 compartment of the blue shelf visible on the tabletop and empty/open to view; its ON
desktop support is consistent."},
  "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the
right compartment of the blue shelf visible on the tabletop; its ON desktop support is
consistent."},
  "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the
middle compartment of the blue shelf visible on the tabletop; its ON desktop support is
consistent."},
  "lipstick": {"status": "unobservable", "evidence": "No complete lipstick assembly is
clearly identifiable in AFTER views; only the lipstick tube/body on the desk is visible, so
the whole lipstick object's current location cannot be established."},
  "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not clearly show a
separate lipstick cap or inside the relevant drawer; its current location remains unverified
."},
  "lipstick_tube": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
show the gold lipstick tube/body standing on the tabletop near the eraser, matching belief
ON desktop."},
  "waste_paper": {"status": "correction_needed", "evidence": "cam_right_wrist AFTER clearly
shows the crumpled paper grasped in the robot gripper rather than resting on the tabletop as
 believed."},
  "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
drawer cabinet on the tabletop with its frame visible; static cabinet presence is consistent
 where observable."},
  "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
 slot and shows its front flush closed, agreeing with belief open=false."},
  "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
drawer slot and shows its front flush closed, agreeing with belief open=false."},
  "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
drawer slot and shows its front flush closed, agreeing with belief open=false."},
  "desktop": {"status": "consistent", "evidence": "cam_high AFTER shows the white tabletop
supporting the visible desk objects; no contradiction to the surface belief."},
  "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the white/
black-rim trash bin on the tabletop at the back-right, matching belief ON desktop."},
  "eraser": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER show the
 small beige eraser resting on the tabletop, matching belief ON desktop."},
  "red_pen": {"status": "unobservable", "evidence": "AFTER views do not show the contents of
 the second drawer or a clearly identifiable red pen elsewhere."},
  "blue_pen": {"status": "unobservable", "evidence": "AFTER views do not show the contents
of the third drawer or a clearly identifiable blue pen elsewhere."},
  "pencil_case": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
show the purple pencil case on the tabletop with lid open, matching belief open=true and ON
desktop."},
  "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the black/
white pen lying inside the open purple pencil case, matching belief INSIDE pencil_case."}
},
"reason": "AFTER images show the drawer cabinet closed at all three layers, the shelf and
desk objects largely unchanged, the previously held book no longer in the gripper with no
unambiguous final book support visible, and the crumpled paper currently held by the gripper
. Only directly visible, unambiguous mismatches are reported.",
"observed_state_delta": [
  {"operation": "observe_held", "node_id": "waste_paper", "confidence": 0.94, "evidence": "
cam_right_wrist AFTER shows the crumpled white paper with blue markings clamped between the
black gripper fingers."},
  {"operation": "disappear", "node_id": "book_1", "confidence": 0.86, "evidence": "
cam_right_wrist BEFORE showed the pending-action book held upright in the gripper;
cam_right_wrist AFTER shows that same gripper no longer holding the book and instead holding
```

```
   crumpled paper, while no clear resting location for the released book is visible in the
  AFTER views."}
  ]
}
```


F.5 STEP 5

Action: `putin(waste_paper,desktop_trash_bin)`

Logged action ID: `7a992c1a77884ae99bc8607573d26ab7`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The waste paper is currently held, and the desktop trash bin is visible. The goal
   requires the waste paper to be placed inside the desktop trash bin, so complete this
   placement first.",
  "action": {"name": "putin", "object": "waste_paper", "target": "desktop_trash_bin", "
   node_ids": ["waste_paper", "desktop_trash_bin"]}
}
```


Evaluator JSON.


```
{"status": "success", "reason": "The crumpled waste paper was moved from the tabletop/gripper
   area and is now clearly inside the desktop trash bin in the after images."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": true,
  "observation_patch_committed": false,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 15.0, "best_cost": 15.0, "stagnant_steps": 0, "
   max_stagnant_steps": 10, "last_cost_change_step": 5, "enabled": true, "cost_before": 16.0, "
   cost_after": 15.0, "cost_changed": true, "improved_historical_best": true},
  "termination_triggered": false,
  "reconciliation_status": "ignored",
  "validation_errors": []
}
```

Step 5: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.93, "identity_evidence": "
   cam_high AFTER shows the white drawer cabinet on the left; counting fixed slots from the top
    identifies the top front as layer 1 above the second and third fronts.", "state_evidence":
   "cam_high AFTER shows the top drawer front seated flush with the cabinet face, not pulled
   out."},
    "drawer_2": {"observed_state": "closed", "confidence": 0.93, "identity_evidence": "
   cam_high AFTER shows the same cabinet with the second front directly below the top front and
    above the bottom front, identifying layer 2.", "state_evidence": "cam_high AFTER shows the
   second drawer front/handle flush in the stack with no exposed tray."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
   cam_high AFTER shows the bottom front of the three-slot white cabinet below the first and
   second fronts, identifying layer 3.", "state_evidence": "cam_high AFTER shows the bottom
   drawer front seated flush; no tray is pulled out."}
   },
   "object_observations": {
    "book_1": {"status": "unobservable", "evidence": "Only one distinct book is clearly
   visible in cam_high AFTER on the desktop; no second book is established in any AFTER view,
   so this bookkeeping book remains unlocalized/lost as believed."},
    "book_2": {"status": "consistent", "evidence": "cam_high AFTER shows a single closed book
   resting ON the desktop near the lower-right area, matching the believed ON desktop placement
    for one interchangeable book slot."},
    "bookshelf": {"status": "unobservable", "evidence": "cam_high AFTER shows the blue multi-
   compartment book rack on the desktop, but the parent furniture node has no specific current
   placement to verify; only its presence is visible."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    book-rack compartments standing on the desktop; the shelf section placement on desktop is
   consistent where visible."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    book-rack compartments standing on the desktop; the third shelf section placement on
   desktop is consistent where visible."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    book-rack compartments standing on the desktop; the second shelf section placement on
   desktop is consistent where visible."},
    "lipstick": {"status": "unobservable", "evidence": "No assembled lipstick object is
   identifiable in AFTER views; visible lipstick-related item on the desk appears to be the
   tube/body, not the complete lipstick object."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not show the
   separate lipstick cap or the inside of the top drawer; its hidden/current location cannot be
    verified."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
   show the lipstick tube/body upright on the desktop near the eraser, matching ON desktop."},
    "waste_paper": {"status": "consistent", "evidence": "cam_right_wrist AFTER clearly shows
   the crumpled waste paper inside the cream desktop trash bin; cam_high AFTER also shows paper
    at the bin rim, consistent with INSIDE desktop_trash_bin."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   drawer cabinet stationary on the desktop; no conflicting visual state for the cabinet as
   furniture."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
    slot and shows its front flush closed, matching the believed closed state."},
    "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
   drawer slot and shows its front flush closed, matching the believed closed state."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
   drawer slot and shows its front flush closed, matching the believed closed state."},
    "desktop": {"status": "consistent", "evidence": "All AFTER cameras show the white desktop
   surface supporting the visible objects; no mismatch with the surface node."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the cream
   trash bin sitting on the desktop at the right side, matching ON desktop; cam_right_wrist
   AFTER shows its open interior."},
    "eraser": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER show the
    small rectangular eraser resting directly on the desktop near the pencil box and lipstick
   body, matching ON desktop."},
    "red_pen": {"status": "unobservable", "evidence": "The second drawer interior is not
   visible in AFTER views and no red pen is identifiable elsewhere, so its hidden location
   cannot be checked."},
    "blue_pen": {"status": "unobservable", "evidence": "The third drawer interior is not
   visible in AFTER views and no blue pen is identifiable elsewhere, so its hidden location
   cannot be checked."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
   show the purple pencil box on the desktop with its lid open, matching believed ON desktop
   and open state."},
    "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the black
   pen lying inside the open purple pencil box, matching INSIDE pencil_case."}
   },
   "reason": "AFTER views confirm the requested waste paper placement in the desktop trash bin
   and show the drawer stack closed. Other visible desktop objects match their believed
   supports; hidden drawer contents and unobserved lost/occluded objects are not inferred.",
```

```
  "observed_state_delta": []
}
```


F.6 STEP 6

Action: `open(drawer_2)`

Logged action ID: `d33f3ea7c0944278814fffb61d293f63`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The red pen is still not visible, and the goal requires it to be placed in the
   pencil case. Open the visible drawer_2 first to search for and reveal the red pen.",
  "action": {"name": "open", "object": "drawer_2", "node_ids": ["drawer_2"]}
}
```


Evaluator JSON.


```
{"status": "failure", "reason": "The second drawer of the white drawer unit remains closed
   with no usable opening or exposed interior visible after the action."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": false,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 17.0, "best_cost": 15.0, "stagnant_steps": 0, "
   max_stagnant_steps": 10, "last_cost_change_step": 6, "enabled": true, "cost_before": 15.0, "
   cost_after": 17.0, "cost_changed": true, "improved_historical_best": false},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```


Step 6: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
```

```
  "drawer_1": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
cam_high AFTER shows the white drawer cabinet on the left; counting fixed slots from the top
, the top drawer front is the first layer above the second and third fronts.", "
state_evidence": "cam_high AFTER shows the top drawer front seated flush with the cabinet
face and no tray protruding."},
  "drawer_2": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
 AFTER shows the same cabinet; the middle visible front between the top and bottom fronts is
 the second layer.", "state_evidence": "cam_high AFTER shows the middle drawer front aligned
 flush with the cabinet, with no pulled-out tray visible."},
  "drawer_3": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
 AFTER shows the bottom front of the three-slot white cabinet, below the first and second
fronts, identifying the third layer.", "state_evidence": "cam_high AFTER shows the bottom
drawer front seated flush, with no outward displacement."}
},
"object_observations": {
  "book_1": {"status": "unobservable", "evidence": "Only one distinct book is clearly
visible on the desktop in cam_high AFTER; the interchangeable book group permits assigning
that visible book to one slot, but no second book is established in the AFTER views."},
  "book_2": {"status": "consistent", "evidence": "cam_high AFTER shows one book resting flat
 ON the desktop near the lower right, matching the belief for one interchangeable book on
desktop; it appears closed."},
  "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue upright
book rack on the tabletop; as furniture with no tracked placement, no contradiction is
visible."},
  "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
 rack standing on the desktop with its compartments visible; the first compartment remains
part of the rack on the desktop, with no conflicting state observed."},
  "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
 rack on the desktop; the third compartment is visible as part of the rack and no
contradictory placement/state is observed."},
  "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
 rack on the desktop; the middle compartment remains visible as part of the rack and no
conflicting placement/state is observed."},
  "lipstick": {"status": "unobservable", "evidence": "No complete assembled lipstick is
clearly identifiable in the AFTER views; only the separate lipstick tube/body on the desk is
 visible."},
  "lipstick_cap": {"status": "unobservable", "evidence": "The lipstick cap is not clearly
visible in the AFTER views; drawer interiors are closed/unobservable, so its current
location cannot be checked."},
  "lipstick_tube": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
AFTER show the lipstick tube/body standing on the desktop near the eraser, matching ON
desktop."},
  "waste_paper": {"status": "correction_needed", "evidence": "cam_high AFTER shows a
crumpled paper now lying on the desktop beside the book outside the trash bin, while
cam_right_wrist AFTER still shows paper inside the bin; the catalog waste_paper belief says
INSIDE desktop_trash_bin, but the clearly moved crumpled paper is ON desktop."},
  "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
drawer cabinet on the desktop; no tracked movable placement/state for the furniture itself
is contradicted."},
  "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
 front and shows it closed/flush, matching the believed closed state."},
  "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
drawer front and shows it closed/flush, matching the believed closed state."},
  "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
drawer front and shows it closed/flush, matching the believed closed state despite the
failed open action."},
  "desktop": {"status": "consistent", "evidence": "cam_high AFTER shows the tabletop surface
 visible and supporting the visible objects; no contradictory state applies."},
  "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER and
cam_right_wrist AFTER show the cream trash bin on the desktop; it remains ON desktop, though
 not all contents match the waste_paper belief."},
  "eraser": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist AFTER
show the small beige eraser resting on the desktop, matching ON desktop."},
  "red_pen": {"status": "unobservable", "evidence": "The red pen is not visible in the AFTER
 views, and the second drawer is observed closed, so its hidden location cannot be directly
checked."},
  "blue_pen": {"status": "unobservable", "evidence": "The blue pen is not visible in the
AFTER views, and the third drawer is observed closed, so its hidden location cannot be
directly checked."},
  "pencil_case": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
AFTER show the purple pencil box open on the desktop, matching the believed open state and
ON desktop."},
  "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the black
pen inside the open purple pencil box, matching INSIDE pencil_case."}
},
"reason": "AFTER images show the drawer cabinet fronts flush closed for all three fixed
layers. Most visible desktop objects agree with current belief. The only clear mismatch is a
 crumpled paper visible on the desktop outside the trash bin in cam_high AFTER, while the
catalog paper was believed inside the desktop trash bin. Hidden drawer contents and the
second book are not established by direct AFTER evidence.",
"observed_state_delta": [
```

```
    {"operation": "relocate", "node_id": "waste_paper", "relation": "ON", "target": "desktop",
    "confidence": 0.86, "evidence": "cam_high AFTER shows the crumpled paper resting on the
   tabletop to the right of the book and outside the cream trash bin; its direct support is the
    desktop, not the bin."}
  ]
}
```


F.7 STEP 7

Action: `grab(waste_paper)`

Logged action ID: `bbc78898773d4f8c9335fbcdfdfa1d20`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The waste paper is currently visible on the desktop, and the goal requires it to
   be placed in the desktop trash bin. Both hands are free, so pick up the waste paper first.",
  "action": {"name": "grab", "object": "waste_paper", "node_ids": ["waste_paper"]}
}
```


Evaluator JSON.


```
{"status": "failure", "reason": "The waste paper remains in the same visible position near the
   trash bin in the before and after views; it has not been grasped or lifted."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": false,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 15.0, "best_cost": 15.0, "stagnant_steps": 0, "
   max_stagnant_steps": 10, "last_cost_change_step": 7, "enabled": true, "cost_before": 17.0, "
   cost_after": 15.0, "cost_changed": true, "improved_historical_best": false},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```

Step 7: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.93, "identity_evidence": "
   cam_high AFTER shows the white three-drawer cabinet from above; the top drawer front is the
   first slot directly under the open top panel, above the two lower fronts.", "state_evidence
   ": "In cam_high AFTER the top drawer front appears seated flush with the cabinet face; no
   tray is pulled out."},
    "drawer_2": {"observed_state": "closed", "confidence": 0.93, "identity_evidence": "
   cam_high AFTER shows the same cabinet; the second layer is the middle drawer front between
   the top and bottom fronts.", "state_evidence": "In cam_high AFTER the middle drawer front is
    flush with the cabinet face; no displacement or exposed tray is visible."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.93, "identity_evidence": "
   cam_high AFTER shows the bottom drawer front as the third slot below the top and middle
   fronts.", "state_evidence": "In cam_high AFTER the bottom drawer front is seated flush; no
   tray is pulled out."}
   },
   "object_observations": {
    "book_1": {"status": "unobservable", "evidence": "Books are interchangeable, but only one
   distinct book is clearly visible on the desk in cam_high AFTER; there is no second distinct
   book localized for this bookkeeping slot."},
    "book_2": {"status": "consistent", "evidence": "cam_high AFTER shows one closed book lying
    on the desktop near the front-right area, matching a book ON desktop under interchangeable
   assignment."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue book rack
    standing on the desktop near the rear center; its static support is visually unchanged."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    rack compartments on the desktop; the first compartment is visible as part of the rack and
   agrees with being on the desk."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    rack compartments on the desktop; the third compartment is visible as part of the rack and
   agrees with being on the desk."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    rack compartments on the desktop; the second compartment is visible as part of the rack and
    agrees with being on the desk."},
    "lipstick": {"status": "unobservable", "evidence": "No assembled lipstick object is
   directly identifiable in AFTER views; only a lipstick-like tube component on the desk is
   visible, so this object's location cannot be checked."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not show an
   identifiable lipstick cap; drawer interiors are not visible because all drawer fronts appear
    closed."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
   show the lipstick tube/body upright on the desktop near the eraser, consistent with ON
   desktop."},
    "waste_paper": {"status": "correction_needed", "evidence": "cam_right_wrist AFTER clearly
   shows the crumpled waste paper resting inside the desktop trash bin, not on the desktop as
   believed."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   drawer cabinet stationary on the desktop; its numbered fronts are visible."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
    slot and shows its front flush closed, matching the believed closed state."},
    "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
   drawer slot and shows its front flush closed, matching the believed closed state."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
   drawer slot and shows its front flush closed, matching the believed closed state."},
    "desktop": {"status": "consistent", "evidence": "AFTER cameras show the tabletop surface
   supporting the drawer cabinet, book rack, trash bin, pencil box, eraser, lipstick body, and
   book."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high and cam_right_wrist
   AFTER show the trash bin on the desktop at the rear-right area, consistent with ON desktop
   ."},
    "eraser": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER show the
    small rectangular eraser resting on the desktop near the center-left, consistent with ON
   desktop."},
    "red_pen": {"status": "unobservable", "evidence": "No red pen is visible in AFTER views;
   drawer interiors are not visible with the drawer fronts closed, so its hidden location
   cannot be checked."},
    "blue_pen": {"status": "unobservable", "evidence": "No blue pen is visible in AFTER views;
    the third drawer interior is not visible because the drawer appears closed."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
   show the purple pencil case open on the desktop, matching its believed open state and ON
   desktop placement."},
    "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the black
   pen inside the open purple pencil case, consistent with INSIDE pencil_case."}
   },
   "reason": "AFTER images show the attempted grab did not result in the paper being held;
   instead the crumpled paper is directly visible inside the desktop trash bin. The drawer
   cabinet is visible as three fixed layers, all closed. Other directly visible objects match
```

```
  their believed placements; hidden or absent items are marked unobservable rather than
  inferred.",
  "observed_state_delta": [
    {"operation": "relocate", "node_id": "waste_paper", "relation": "INSIDE", "target": "
  desktop_trash_bin", "confidence": 0.96, "evidence": "cam_right_wrist AFTER provides a close
  view into the desktop trash bin with the crumpled waste paper clearly inside the bin;
  cam_high AFTER also shows the paper mass at the bin area rather than on the open desktop."}
  ]
}
```


F.8 STEP 8

Action: `grab(eraser)`

Logged action ID: `b3534a95349d4057a954acf0289b841e`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The eraser is visible on the desktop, and the pencil case is open. The goal
  requires the eraser to be inside the pencil case. Both hands are currently free, so pick up
  the eraser first.",
  "action": {"name": "grab", "object": "eraser", "node_ids": ["eraser"]}
}
```


Evaluator JSON.


```
{"status": "failure", "reason": "The eraser remains on the table in the same position before
  and after; it is not visibly grasped or lifted by the gripper."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": false,
  "observation_patch_committed": false,
  "graph_update_committed": false,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 15.0, "best_cost": 15.0, "stagnant_steps": 1, "
  max_stagnant_steps": 10, "last_cost_change_step": 7, "enabled": true, "cost_before": 15.0, "
  cost_after": 15.0, "cost_changed": false, "improved_historical_best": false},
  "termination_triggered": false,
  "reconciliation_status": "ignored",
  "validation_errors": []
}
```

Step 8: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.93, "identity_evidence": "
   cam_high AFTER shows the white drawer cabinet from above-left with three fixed fronts; the
   top slot/front is identifiable directly under the cabinet top panel.", "state_evidence": "
   cam_high AFTER shows the top drawer front seated flush with the cabinet, with no tray
   projecting."},
    "drawer_2": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
   cam_high AFTER shows the same cabinet with three vertically stacked fronts; the second slot
   is the middle front between the top and bottom handles.", "state_evidence": "cam_high AFTER
   shows the middle drawer front seated flush, not pulled out."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
   cam_high AFTER shows the bottom slot/front of the three-layer cabinet below the middle
   drawer.", "state_evidence": "cam_high AFTER shows the bottom drawer front seated flush with
   no visible open tray."}
   },
   "object_observations": {
    "book_1": {"status": "unobservable", "evidence": "cam_high AFTER shows one book on the
   desk, but the interchangeable book group has one lost slot and one believed-on-desk slot;
   only one physical book is established, assigned to book_2, so book_1 is not separately
   observable."},
    "book_2": {"status": "consistent", "evidence": "cam_high AFTER shows a single book lying
   flat on the desk at the front-right area, consistent with book_2 ON desktop under the
   interchangeable bookkeeping assignment."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue upright
   book rack standing on the desk; static furniture placement is visually unchanged."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    book rack on the desk with its left/top compartment structure visible; its desk support is
   consistent."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    book rack on the desk with the right/third compartment visible; its desk support is
   consistent."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    book rack on the desk with the middle compartment visible; its desk support is consistent
   ."},
    "lipstick": {"status": "unobservable", "evidence": "No assembled lipstick object is
   clearly visible in AFTER views; only a lipstick-like tube/body on the desk can be checked
   separately as lipstick_tube."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not clearly show a
   separate lipstick cap or the inside of the top drawer where it is believed occluded."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
   show the lipstick tube/body standing upright on the desk near the eraser, consistent with ON
    desktop."},
    "waste_paper": {"status": "consistent", "evidence": "cam_right_wrist AFTER looks into the
   tabletop trash bin and shows crumpled paper inside it, consistent with waste_paper INSIDE
   desktop_trash_bin."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   drawer cabinet on the desktop in the expected area; the cabinet itself is static."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
    layer and shows its front closed/flush, matching the believed closed state."},
    "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
   drawer layer and shows its front closed/flush, matching the believed closed state."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
   drawer layer and shows its front closed/flush, matching the believed closed state."},
    "desktop": {"status": "consistent", "evidence": "All AFTER cameras show the white tabletop
    supporting the visible drawer cabinet, rack, bin, pencil box, eraser, lipstick body, and
   book; no contradiction to static surface belief."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high and cam_right_wrist
   AFTER show the cream trash bin with black rim sitting on the desk, consistent with ON
   desktop."},
    "eraser": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER show the
    small beige eraser still resting on the desk between the pencil box and lipstick body; it
   is not held."},
    "red_pen": {"status": "unobservable", "evidence": "The middle drawer is closed in cam_high
    AFTER and its contents are not visible; no red pen is directly observed elsewhere."},
    "blue_pen": {"status": "unobservable", "evidence": "The bottom drawer is closed in
   cam_high AFTER and its contents are not visible; no blue pen is directly observed elsewhere
   ."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
   show the purple pencil box on the desk with its lid open, matching the believed open state
   ."},
    "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows a black/
   white pen lying inside the open purple pencil box, consistent with black_pen INSIDE
   pencil_case."}
   },
   "reason": "AFTER images show the attempted grab failed: the eraser remains on the desktop
   and the grippers do not hold it. The three drawer layers are clearly identifiable in
```

```
  cam_high AFTER and all are closed. Other visible objects match their believed placements;
  hidden drawer contents and the extra lost book are not directly observable.",
  "observed_state_delta": []
}
```


F.9 STEP 12

Action: `open(drawer_2)`

Logged action ID: `26e8822c7dfd4571ab736a6eeb040556`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The red and blue pens are still not visible, and the goal requires them to be
  placed in the pencil case. Open the visible drawer_2 first to attempt to reveal the red pen
  .",
  "action": {"name": "open", "object": "drawer_2", "node_ids": ["drawer_2"]}
}
```


Evaluator JSON.


```
{"status": "failure", "reason": "The second drawer is not visibly open; after images show the
  drawer unit still closed/unchanged, while the robot appears to have moved a book instead."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": false,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 14.0, "best_cost": 14.0, "stagnant_steps": 0, "
  max_stagnant_steps": 10, "last_cost_change_step": 12, "enabled": true, "cost_before": 15.0,
  "cost_after": 14.0, "cost_changed": true, "improved_historical_best": true},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```

Step 12: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "
   cam_left_wrist AFTER shows the white drawer cabinet with three fixed fronts counted from the
    top; the uppermost slot/front is visible above the second and third handles.", "
   state_evidence": "The top drawer front appears seated flush with the cabinet face; no tray
   is pulled out."},
    "drawer_2": {"observed_state": "closed", "confidence": 0.88, "identity_evidence": "
   cam_left_wrist AFTER shows the same cabinet; the second slot is the middle front between the
    top and bottom drawer handles.", "state_evidence": "The middle drawer front appears flush
   with the frame and not displaced outward; no middle tray is exposed."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.88, "identity_evidence": "
   cam_left_wrist AFTER shows the bottom slot below the first and second fronts in the fixed
   three-layer cabinet.", "state_evidence": "The bottom drawer front appears seated flush with
   the cabinet face; no bottom tray is pulled out."}
   },
   "object_observations": {
    "book_1": {"status": "correction_needed", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show one book clearly upright and held by the gripper near the drawer/bookshelf area,
   not resting ON the desk as believed."},
    "book_2": {"status": "unobservable", "evidence": "Only one distinct book is clearly
   established in AFTER views; the other interchangeable book cannot be separately localized or
    confirmed on the desk."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue file/
   bookshelf standing on the desktop in its expected scene position; no relocation of this
   static furniture is indicated."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    shelf compartments upright on the desk; the first compartment is visible as part of the
   shelf, with no contradictory placement observed."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    shelf compartments upright on the desk; the third compartment is visible as part of the
   shelf, with no contradictory placement observed."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    shelf compartments upright on the desk; the middle compartment is visible as part of the
   shelf, with no contradictory placement observed."},
    "lipstick": {"status": "unobservable", "evidence": "No complete lipstick assembly is
   identifiable in the AFTER views; only a small upright lipstick-like item/part near the desk
   center is partly visible, insufficient to bind to this aggregate ID."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not clearly show a
   separate lipstick cap or the inside of the top drawer where it may be occluded."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high AFTER shows the small
   upright lipstick body on the desktop near the center/front area, matching the belief that it
    is ON the desk."},
    "waste_paper": {"status": "consistent", "evidence": "cam_high AFTER shows crumpled white
   paper inside the cream desktop trash bin, matching INSIDE desktop_trash_bin."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show the white three-layer drawer cabinet stationary on the desk; no relocation or
   state property for the cabinet itself is contradicted."},
    "drawer_1": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the top
   drawer front seated flush and closed, matching the post-action belief open=false."},
    "drawer_3": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the bottom
   drawer front seated flush and closed, matching the post-action belief open=false."},
    "drawer_2": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the middle
   drawer front seated flush and closed despite the attempted open action, matching the failure
   /current belief open=false."},
    "desktop": {"status": "consistent", "evidence": "AFTER views show the white tabletop
   supporting the drawer cabinet, shelf, trash bin, lipstick body, eraser, and pencil box area;
    no contradiction for the static surface."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the cream
   trash bin on the desktop with crumpled paper inside, matching its believed placement ON
   desktop."},
    "eraser": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the small beige
    eraser resting on the desktop near the lower-left/center area, matching ON desktop."},
    "red_pen": {"status": "unobservable", "evidence": "The second drawer is observed closed in
    cam_left_wrist AFTER, and no red pen is directly visible elsewhere; hidden contents cannot
   be checked."},
    "blue_pen": {"status": "unobservable", "evidence": "The third drawer is observed closed in
    cam_left_wrist AFTER, and no blue pen is directly visible elsewhere; hidden contents cannot
    be checked."},
    "pencil_case": {"status": "unobservable", "evidence": "The purple pencil box area is
   mostly out of view/occluded in AFTER images; its open state and exact support cannot be
   directly verified."},
    "black_pen": {"status": "unobservable", "evidence": "AFTER views do not clearly show the
   inside of the pencil box or the black pen; its believed INSIDE pencil_case placement cannot
   be verified."}
   },
   "reason": "AFTER images show the attempted drawer opening did not leave any drawer visibly
   open; all three drawer fronts are flush. The main clear mismatch is that one interchangeable
```

```
    book is now held upright in the gripper rather than lying on the desktop. Other visible
   objects either match their believed support/container or are not sufficiently observable.",
   "observed_state_delta": [
     {"operation": "observe_held", "node_id": "book_1", "confidence": 0.95, "evidence": "
    cam_high AFTER shows the book cover clamped by the gripper on the left side of the scene;
    cam_left_wrist AFTER shows the book spine held vertically between the gripper jaws."}
   ]
}
```


F.10 STEP 13

Action: `putin(book_1,shelf_compartment_1)`

Logged action ID: `1aa6a40cd300445a8d58e764e3b87a96`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "book_1 is currently held, and shelf_compartment_1 is visible. The goal requires
   book_1 to be placed in shelf_compartment_1, so place the held book in shelf_compartment_1
   first.",
   "action": {"name": "putin", "object": "book_1", "target": "shelf_compartment_1", "node_ids":
    ["book_1", "shelf_compartment_1"]}
}
```


Evaluator JSON.


```
{"status": "success", "reason": "After action, the book is visibly placed upright inside/
   against the first compartment of the blue shelf, whereas before it was outside the shelf."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": true,
  "observation_patch_committed": false,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 13.0, "best_cost": 13.0, "stagnant_steps": 0, "
   max_stagnant_steps": 10, "last_cost_change_step": 13, "enabled": true, "cost_before": 14.0,
   "cost_after": 13.0, "cost_changed": true, "improved_historical_best": true},
  "termination_triggered": false,
  "reconciliation_status": "ignored",
  "validation_errors": []
}
```

Step 13: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.94, "identity_evidence": "
   cam_high AFTER shows the white three-slot drawer cabinet at left; the top slot is identified
    by counting the highest front directly below the open top rim and above the two lower
   fronts.", "state_evidence": "cam_high AFTER shows the top drawer front seated flush with the
    cabinet, with no tray pulled outward."},
    "drawer_2": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
   cam_high AFTER shows the same cabinet; the second layer is the middle front between the top
   drawer front and the bottom drawer front.", "state_evidence": "cam_high AFTER shows the
   middle drawer front and handle aligned flush with the cabinet face, not displaced open."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
    AFTER shows the cabinet; the third layer is the lowest front beneath the middle drawer and
   above the tabletop.", "state_evidence": "cam_high AFTER shows the lower drawer front appears
    seated flush; no exposed tray is visible."}
   },
   "object_observations": {
    "book_1": {"status": "consistent", "evidence": "cam_high AFTER and cam_right_wrist AFTER
   show one book placed upright within the blue file rack's left/front compartment area;
   assigning this observed book to interchangeable slot book_1 agrees with belief INSIDE
   shelf_compartment_1 and not held."},
    "book_2": {"status": "correction_needed", "evidence": "cam_high AFTER no longer shows the
   other book standing on the left desktop where it was before; only one distinct book is
   visible in the shelf, and no second book is visible on the desktop. This contradicts belief
   ON desktop, but its current location is not established because the old site/near left edge
   is partly changed and book identity is interchangeable."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue vertical
   file shelf/rack on the tabletop in the same central-right position; static furniture
   placement is visually consistent."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the left
   /front compartment of the blue shelf on the tabletop containing the upright book; the
   compartment remains on the desk and open for insertion."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    shelf with its right-side compartment on the tabletop; no contradiction to its believed ON
   desktop placement is visible."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the
   middle compartment of the blue shelf on the tabletop; no contradictory placement or state is
    observed."},
    "lipstick": {"status": "unobservable", "evidence": "No complete assembled lipstick object
   is clearly visible in the AFTER views; only the small tube/body-like item on the desktop can
    be checked separately as lipstick_tube."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not show a distinct
    lipstick cap; the top drawer is closed, so any believed hidden cap cannot be directly
   checked."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show the small lipstick tube/body upright on the desktop near the center-front area,
   agreeing with belief ON desktop and not held."},
    "waste_paper": {"status": "consistent", "evidence": "cam_high AFTER shows crumpled white
   paper inside the cream tabletop trash bin at the back-right, consistent with belief INSIDE
   desktop_trash_bin."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   three-drawer cabinet on the left side of the tabletop; the furniture itself remains in place
   ."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
    slot of the cabinet and shows it closed/flush, matching belief open=false; placement as
   part of cabinet on tabletop is unchanged."},
    "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
   drawer slot and shows it closed/flush, matching belief open=false."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
   drawer slot and shows it closed/flush, matching belief open=false."},
    "desktop": {"status": "consistent", "evidence": "All AFTER cameras show the white tabletop
    supporting the shelf, drawer cabinet, trash bin, pencil box, eraser, and lipstick body; no
   contradiction to static surface belief."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the cream
   trash bin with black rim on the back-right of the tabletop, with paper inside, matching
   belief ON desktop."},
    "eraser": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist AFTER
   show the small beige eraser lying on the desktop near the front-center/left of the lipstick
   body, matching belief ON desktop."},
    "red_pen": {"status": "unobservable", "evidence": "No red pen is visible in the AFTER
   images; drawer contents are not exposed because all drawer fronts appear closed, so its
   hidden/unknown belief cannot be checked."},
    "blue_pen": {"status": "unobservable", "evidence": "No blue pen is visible in the AFTER
   images; the third drawer is closed, so any hidden location cannot be directly verified."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show the purple pencil box on the left-front tabletop with lid open, consistent with
   belief ON desktop and open=true."},
```

```
    "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows a black/
   white marker-like pen lying inside the open purple pencil box, agreeing with belief INSIDE
   pencil_case."}
   },
   "reason": "AFTER views show the action book now upright in the blue shelf, while the three
   drawer layers are all directly visible as closed in the high camera. Most unrelated objects
   match their believed visible placements. The only visual mismatch is that no second distinct
    book is established on the desktop; however, because books are interchangeable and only one
    book is clearly localized, no unambiguous relocate/disappear operation for book_2 is
   supported under the required evidence standard.",
   "observed_state_delta": []
}
```


F.11 STEP 14

Action: `open(drawer_3)`

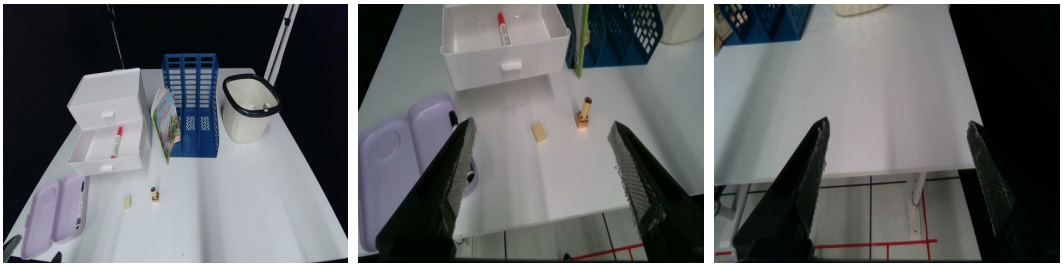
Logged action ID: `4c825a4fa12c4a5ebbca3584f59ef7a4`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The blue pen is still not visible, and the goal requires it to be placed in the
   pencil case. Both hands are currently free, so open the visible drawer_3 first to attempt to
    reveal the blue pen.",
  "action": {"name": "open", "object": "drawer_3", "node_ids": ["drawer_3"]}
}
```


Evaluator JSON.


```
{"status": "success", "reason": "The target third drawer is visibly pulled open far enough to
   expose its interior/access area, with a pen visible inside."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": true,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 13.0, "best_cost": 13.0, "stagnant_steps": 1, "
   max_stagnant_steps": 10, "last_cost_change_step": 13, "enabled": true, "cost_before": 13.0,
    "cost_after": 13.0, "cost_changed": false, "improved_historical_best": false},
  "termination_triggered": false,
```

```
  "reconciliation_status": "success",
  "validation_errors": []
}
```


Step 14: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
   cam_high AFTER shows the white drawer cabinet with three fixed slots counted from the top;
   the top slot/front is the first layer above the pulled-out middle drawer.", "state_evidence
   ": "cam_high AFTER shows the top drawer front seated flush with the cabinet and no top tray
   protruding."},
    "drawer_2": {"observed_state": "open", "confidence": 0.95, "identity_evidence": "cam_high
   AFTER shows the cabinet top and all drawer fronts; the protruding tray is attached to the
   middle front between the top and bottom fronts, identifying the second layer.", "
   state_evidence": "cam_high AFTER and cam_left_wrist AFTER show the middle drawer pulled
   outward with its tray exposed and a red pen inside."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
    AFTER shows the lowest fixed slot/front below the protruding middle drawer, identifying the
    third layer from top to bottom.", "state_evidence": "cam_high AFTER shows the bottom drawer
    front/handle aligned with the cabinet body rather than pulled out."}
  },
  "object_observations": {
    "book_1": {"status": "consistent", "evidence": "cam_high AFTER shows one book standing
   inside the left compartment/first visible section of the blue shelf; assigned to
   interchangeable book slot book_1, matching INSIDE shelf_compartment_1 and closed appearance
   ."},
    "book_2": {"status": "unobservable", "evidence": "cam_high AFTER does not clearly show a
   separate second book on the tabletop; the visible book/shelf area is partly occluded by the
   blue rack and no distinct second book placement can be confirmed."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue bookshelf
   /rack stationary on the desktop; furniture location is not otherwise relocatable."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the left
    shelf compartment on the desk containing the visible book, consistent with the shelf
   compartment being on the desktop."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    shelf structure on the desk; the third compartment is visible as part of the stationary
   rack, with no contradictory placement observed."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
    shelf structure on the desk; the middle compartment is visible as part of the stationary
   rack, with no contradictory placement observed."},
    "lipstick": {"status": "unobservable", "evidence": "No assembled lipstick object is
   directly identifiable in AFTER views; only the lipstick tube/body on the desk is visible."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not clearly reveal
   a separate lipstick cap or the inside of the top drawer where it was believed occluded; no
   location can be established."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show the lipstick tube/body upright on the desktop near the eraser, consistent with ON
    desktop."},
    "waste_paper": {"status": "consistent", "evidence": "cam_high AFTER shows crumpled white
   paper inside the white desktop trash bin at the rear right, consistent with INSIDE
   desktop_trash_bin."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   drawer cabinet stationary on the desktop; the cabinet itself is not relocatable."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top layer
   front as closed/flush, matching belief open=false; it remains part of the cabinet on the
   desktop."},
    "drawer_3": {"status": "correction_needed", "evidence": "cam_high AFTER identifies the
   bottom layer as closed/flush, contradicting belief open=true for drawer_3."},
    "drawer_2": {"status": "correction_needed", "evidence": "cam_high AFTER and cam_left_wrist
    AFTER identify the middle layer as pulled open with tray exposed, contradicting belief open
   =false for drawer_2."},
    "desktop": {"status": "consistent", "evidence": "AFTER views show the white tabletop
   supporting the drawer cabinet, shelf, trash bin, pencil box, eraser, and lipstick body; no
   contradictory state."},
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   trash bin with black rim on the desktop, with paper inside, consistent with ON desktop."},
    "eraser": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist AFTER
   show the small beige eraser resting on the desktop near the lipstick body, consistent with
   ON desktop."},
    "red_pen": {"status": "correction_needed", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER clearly show a red-capped pen lying inside the open middle drawer, whereas belief has
   no location for red_pen."},
    "blue_pen": {"status": "unobservable", "evidence": "The third/bottom drawer is closed in
   cam_high AFTER and no blue pen is clearly visible elsewhere; its current placement cannot be
    checked."},
```

```
    "pencil_case": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist
   AFTER show the purple pencil box open on the desktop, consistent with open=true and ON
   desktop."},
    "black_pen": {"status": "consistent", "evidence": "cam_high AFTER and cam_left_wrist AFTER
    show the black pen inside the open purple pencil box, consistent with INSIDE pencil_case."}
   },
   "reason": "AFTER images show the action actually left the middle drawer (second layer) open,
    not the bottom/third layer. The top and bottom drawer fronts are flush closed. The red pen
   is directly visible inside the open middle drawer, while other visible desktop and shelf
   objects largely match belief; several hidden or ambiguous objects cannot be localized from
   the AFTER views.",
   "observed_state_delta": [
    {"operation": "relocate", "node_id": "red_pen", "relation": "INSIDE", "target": "drawer_2
   ", "confidence": 0.93, "evidence": "cam_high AFTER and cam_left_wrist AFTER show the red-
   capped pen lying in the exposed tray of the middle drawer, identified as drawer_2 by its
   position between the top and bottom fronts."},
    {"operation": "reveal", "source": "drawer_2", "node_id": "red_pen", "confidence": 0.91, "
   evidence": "cam_high AFTER shows drawer_2 open with the previously occluded red pen now
   visible inside its tray."}
   ]
}
```

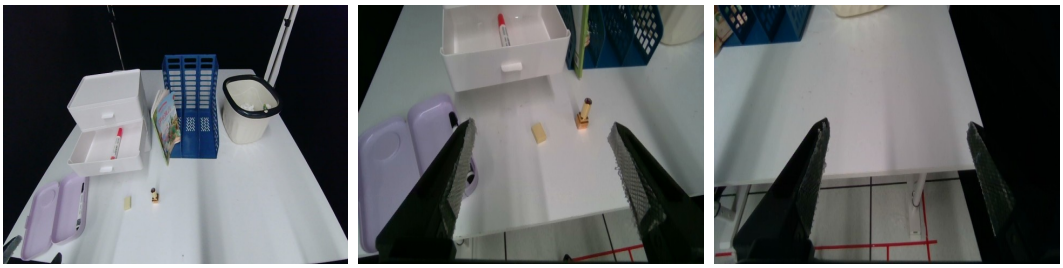

F.12 STEP 15

Action: `grab(red_pen)`

Logged action ID: `55694239feae45b0be59ce1e688d4652`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The red pen is now visible inside the open drawer_2. The goal requires it to be
   placed in the open pencil case. Both hands are currently free, so pick up the red pen first
   .",
  "action": {"name": "grab", "object": "red_pen", "node_ids": ["red_pen"]}
}
```


Evaluator JSON.


```
{"status": "success", "reason": "After action, the red pen is clearly held between the gripper
    jaws, whereas before it was lying in the tray/drawer area."}
```

Execution record (selected backend fields).


```
{
  "action_effect_committed": true,
  "observation_patch_committed": false,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 12.0, "best_cost": 12.0, "stagnant_steps": 0, "
  max_stagnant_steps": 10, "last_cost_change_step": 15, "enabled": true, "cost_before": 13.0,
  "cost_after": 12.0, "cost_changed": true, "improved_historical_best": true},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```


Step 15: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.88, "identity_evidence": "
  cam_high AFTER shows the white drawer cabinet at left with fixed top/bottom frame; the top
  slot/front is the layer immediately under the cabinet top, above the pulled-out middle tray
  .", "state_evidence": "cam_high AFTER shows the top drawer front seated flush behind/above
  the open middle tray, with no top tray protruding."},
    "drawer_2": {"observed_state": "open", "confidence": 0.95, "identity_evidence": "cam_high
  AFTER shows the same white cabinet; counting slots from the cabinet top, the visible
  protruding tray with front handle is the second slot, between the flush top drawer and the
  lower area.", "state_evidence": "cam_high AFTER and cam_left_wrist AFTER clearly show the
  second-layer tray pulled forward/open with its interior exposed."},
    "drawer_3": {"observed_state": "unobservable", "confidence": 0.45, "identity_evidence": "
  cam_high AFTER includes the cabinet, but the lowest slot/front is partly hidden by the
  protruding second-layer drawer and camera angle, making the third-layer front difficult to
  distinguish from the cabinet base.", "state_evidence": "No unobstructed AFTER view clearly
  shows the third drawer front seated flush or pulled out."}
  },
  "object_observations": {
    "book_1": {"status": "consistent", "evidence": "cam_high AFTER shows one book assigned to
  book_1 standing inside the left/top compartment area of the blue bookshelf, matching INSIDE
  shelf_compartment_1 under interchangeable book bookkeeping."},
    "book_2": {"status": "consistent", "evidence": "cam_high AFTER shows another book/booklet
  upright on the tabletop immediately left of the blue bookshelf, matching the belief that one
   book is ON desktop under interchangeable book bookkeeping."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue bookshelf
   stationary on the desk; its overall furniture placement is unchanged and no relocation is
  indicated."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the
  first shelf compartment of the blue bookshelf visible on the tabletop with a book inside;
  placement of the compartment/furniture agrees with belief."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the
  rightmost/third blue bookshelf compartment visible and empty on the tabletop; its support on
   the desk is unchanged."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the
  middle blue bookshelf compartment visible on the tabletop; no mismatch with its believed
  desk placement is apparent."},
    "lipstick": {"status": "unobservable", "evidence": "cam_high and wrist AFTER views show
  the lipstick tube/body on the desk, but the assembled whole lipstick as a separate object is
   not established; no clear held or container placement can be verified."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not clearly
  identify a separate lipstick cap; drawer first layer remains closed, so its believed hidden/
  unknown placement cannot be checked visually."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
  show the gold lipstick body standing upright on the desktop near the drawer/bookshelf,
  matching ON desktop."},
    "waste_paper": {"status": "consistent", "evidence": "cam_high AFTER shows crumpled paper
  inside the white tabletop trash bin at the back right, matching INSIDE desktop_trash_bin."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
  drawer cabinet stationary on the tabletop; cabinet identity and placement are unchanged."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
   slot and shows its front flush/closed, matching belief open=false."},
    "drawer_3": {"status": "unobservable", "evidence": "The third/lower drawer front is not
  clearly visible in an unobstructed AFTER view because the open second drawer and angle
  obscure the lower slot; state cannot be confirmed."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
  identify the second slot as the protruding open tray, matching belief open=true."},
    "desktop": {"status": "consistent", "evidence": "AFTER cameras show the white tabletop
  supporting the drawer cabinet, bookshelf, pencil box, eraser, lipstick body and trash bin;
  no mismatch in the surface itself is observed."},
```

```
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   trash bin with black rim resting on the tabletop at the rear right, consistent with ON
   desktop."},
    "eraser": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER show the
    small beige eraser resting on the tabletop near the center-left, matching ON desktop."},
    "red_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the red pen
   clearly clamped in the gripper; cam_high AFTER also shows the pen extending from the robot
   at the left edge, matching held=true."},
    "blue_pen": {"status": "unobservable", "evidence": "No AFTER view clearly shows a blue pen
   ; the believed hidden/unknown location cannot be verified, and the third drawer is not open/
   visible enough to inspect contents."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
   show the purple pencil box open on the tabletop, matching open=true and ON desktop."},
    "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the black
   pen lying inside the open purple pencil box, matching INSIDE pencil_case."}
   },
   "reason": "AFTER images were audited across the full scene. The red pen is visibly held by
   the gripper, the second drawer remains open, the top drawer is closed, and the third drawer
   state is not confidently observable. Other visible objects match their believed placements;
   hidden or not clearly identifiable items are marked unobservable without inferred changes.",
   "observed_state_delta": [
    {"operation": "observe_held", "node_id": "red_pen", "confidence": 0.98, "evidence": "
   cam_left_wrist AFTER shows the red pen centered and clamped between the gripper jaws;
   cam_high AFTER also shows the red-tipped pen protruding from the robot at the left edge."}
   ]
}
```


F.13 STEP 16

Action: `putin(red_pen,pencil_case)`

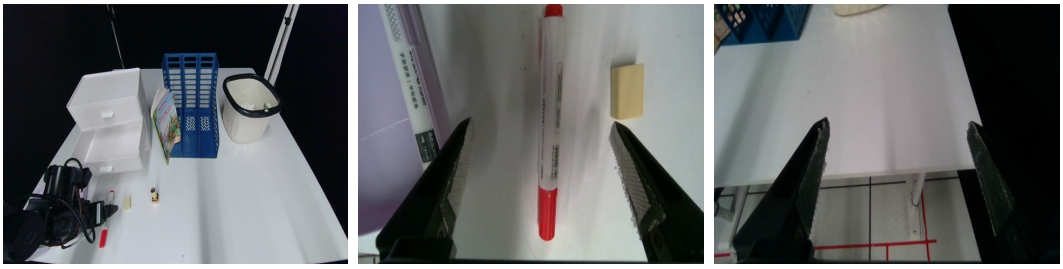
Logged action ID: `49cc12179ede4ad39d47d037a422c17c`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The red pen is currently held, and the pencil case is visible and open. The goal
   requires the red pen to be inside the pencil case, so place the held red pen there first.",
   "action": {"name": "putin", "object": "red_pen", "target": "pencil_case", "node_ids": ["
   red_pen", "pencil_case"]}
}
```


Evaluator JSON.

```
{"status": "failure", "reason": "The red pen is not in the pencil case after the action; it is
    lying on the tabletop next to the open purple pencil case."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": false,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 13.0, "best_cost": 12.0, "stagnant_steps": 0, "
  max_stagnant_steps": 10, "last_cost_change_step": 16, "enabled": true, "cost_before": 12.0,
  "cost_after": 13.0, "cost_changed": true, "improved_historical_best": false},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```


Step 16: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.92, "identity_evidence": "
  cam_high AFTER shows the white drawer cabinet at left with the top slot directly under the
  cabinet top panel, above the middle and bottom drawer fronts.", "state_evidence": "cam_high
  AFTER shows the top drawer front seated flush with the cabinet, with no tray protruding."},
    "drawer_2": {"observed_state": "open", "confidence": 0.93, "identity_evidence": "cam_high
  AFTER shows the second slot from the top, between the top front and the bottom front, in the
   fixed white drawer cabinet.", "state_evidence": "cam_high AFTER shows the middle drawer
  tray/front pulled outward, leaving an open bin visible."},
    "drawer_3": {"observed_state": "closed", "confidence": 0.9, "identity_evidence": "cam_high
   AFTER shows the bottom slot/front below the open middle drawer in the same fixed cabinet.",
   "state_evidence": "cam_high AFTER shows the bottom drawer front seated flush and not pulled
   out."}
  },
  "object_observations": {
    "book_1": {"status": "consistent", "evidence": "cam_high AFTER shows one distinct book
  standing inside/leaning in the left compartment of the blue shelf; interchangeable
  assignment used for the book believed INSIDE shelf_compartment_1."},
    "book_2": {"status": "unobservable", "evidence": "cam_high AFTER does not show a separate
  second book clearly on the desktop; the visible book at the shelf is already assigned to
  book_1, and any other book is not established."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue book rack
   on the desktop in the expected area; it is static furniture with no relocatable state to
  correct."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
   shelf/rack compartment on the desktop holding the visible book, consistent with the
  compartment being on the desk and usable as a container."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
   shelf/rack on the desktop; the right/third compartment area is visible as part of the rack,
   with no contradictory state."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
   shelf/rack on the desktop; the middle compartment area is visible as part of the rack, with
   no contradictory state."},
    "lipstick": {"status": "unobservable", "evidence": "No complete lipstick assembly is
  identifiable in AFTER views; only the tube/body-like item on the desk is visible, so the
  whole lipstick object's current location cannot be checked."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not clearly show an
   identifiable lipstick cap; the closed top drawer may still occlude its believed hidden
  location."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high AFTER shows the lipstick
  tube/body upright on the desktop near the eraser, matching ON desktop."},
    "waste_paper": {"status": "consistent", "evidence": "cam_high AFTER shows crumpled white
  paper inside the tabletop trash bin at the back right, matching INSIDE desktop_trash_bin."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
  drawer cabinet at the left on the desktop; static furniture placement/state has no
  contradictory visual evidence."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
   layer and shows its front flush closed, agreeing with belief open=false."},
    "drawer_3": {"status": "consistent", "evidence": "cam_high AFTER identifies the bottom
  drawer layer and shows its front flush closed, agreeing with belief open=false; contents
  remain obscured."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER identifies the middle
  drawer layer and shows the tray/front pulled outward open, agreeing with belief open=true
  ."},
    "desktop": {"status": "consistent", "evidence": "AFTER views show the white desktop
  supporting the shelf, trash bin, pencil box, eraser, lipstick body, and red pen; no
  contradictory state."},
```

```
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the trash
   bin on the desktop at the back right, with paper inside as believed."},
    "eraser": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER show the
    small yellow eraser resting on the desktop near the red pen, matching ON desktop."},
    "red_pen": {"status": "correction_needed", "evidence": "cam_left_wrist AFTER clearly shows
    the red pen lying on the desktop between the gripper jaws, not held and not inside the
   pencil box; cam_high AFTER also shows it on the desk near the pencil box."},
    "blue_pen": {"status": "unobservable", "evidence": "AFTER views do not show the blue pen;
   the third drawer is closed, so its believed hidden state/location cannot be visually checked
   ."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high AFTER shows the purple
   pencil box on the desktop open at the left edge; cam_left_wrist AFTER also shows the open
   box edge and the black pen inside."},
    "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows the black
   pen lying inside the open purple pencil box, matching INSIDE pencil_case."}
   },
   "reason": "AFTER images show the attempted put-in failed: the red pen is released onto the
   desktop, while the pencil box remains open and the black pen remains inside it. The drawer
   audit is based on cam_high AFTER: top and bottom fronts are flush closed, middle drawer is
   pulled open. Other objects are unchanged where directly visible; hidden or not clearly
   visible objects are marked unobservable rather than inferred.",
   "observed_state_delta": [
    {"operation": "relocate", "node_id": "red_pen", "relation": "ON", "target": "desktop", "
   confidence": 0.96, "evidence": "cam_left_wrist AFTER shows the red pen resting flat on the
   white desktop between open gripper jaws; cam_high AFTER also shows the red pen on the desk
   beside the pencil box, not in the box."}
   ]
}
```

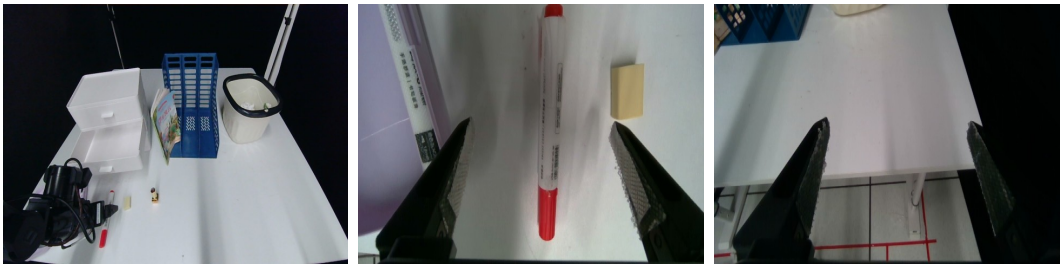

F.14 STEP 17

Action: `grab(red_pen)`

Logged action ID: `2f00726aaa0a456b9355b6968a4cb090`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The red pen is currently visible on the desktop, and the pencil case is open. The
    goal requires the red pen to be inside the pencil case. Both hands are currently free, so
   pick up the red pen first.",
   "action": {"name": "grab", "object": "red_pen", "node_ids": ["red_pen"]}
}
```

Evaluator JSON.


```
{"status": "success", "reason": "The red pen was on the table before and is visibly held
  between the gripper fingers after the action."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": true,
  "observation_patch_committed": true,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 12.0, "best_cost": 12.0, "stagnant_steps": 0, "
  max_stagnant_steps": 10, "last_cost_change_step": 17, "enabled": true, "cost_before": 13.0,
  "cost_after": 12.0, "cost_changed": true, "improved_historical_best": false},
  "termination_triggered": false,
  "reconciliation_status": "success",
  "validation_errors": []
}
```


Step 17: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"observed_state": "closed", "confidence": 0.86, "identity_evidence": "
  cam_high AFTER shows the white drawer cabinet from above/side with the top slot directly
  below the cabinet top and above the protruding middle drawer.", "state_evidence": "The top
  drawer front is seated flush/closed with no tray extended."},
    "drawer_2": {"observed_state": "open", "confidence": 0.92, "identity_evidence": "cam_high
  AFTER shows three fixed drawer slots; the middle front between the top closed drawer and
  lower area is the second layer.", "state_evidence": "The second-layer tray/front is clearly
  pulled outward, exposing its empty tray."},
    "drawer_3": {"observed_state": "unobservable", "confidence": 0.55, "identity_evidence": "
  cam_high AFTER gives only a partial view of the cabinet below the open second drawer; the
  lowest slot is largely obscured by the protruding middle tray and camera angle.", "
  state_evidence": "The third drawer front is not clearly visible enough to determine whether
  it is flush closed or displaced open."}
  },
  "object_observations": {
    "book_1": {"status": "consistent", "evidence": "cam_high AFTER shows a book standing in
  the left/front compartment of the blue shelf; assigned to interchangeable book slot book_1,
  matching INSIDE shelf_compartment_1."},
    "book_2": {"status": "correction_needed", "evidence": "cam_high AFTER shows a second
  upright book/green-backed item in the blue shelf area, not a book resting ON the desktop as
  believed; assign interchangeable slot book_2 to a shelf compartment placement."},
    "bookshelf": {"status": "consistent", "evidence": "cam_high AFTER shows the blue shelf
  standing on the desktop; static furniture placement is visually unchanged."},
    "shelf_compartment_1": {"status": "consistent", "evidence": "cam_high AFTER shows the left
  /front shelf compartment on the desktop and containing an upright book, consistent with
  being available on the desk."},
    "shelf_compartment_3": {"status": "consistent", "evidence": "cam_high AFTER shows the blue
   shelf's right/back compartment structure on the desktop; no contradictory state or
  placement is observed."},
    "shelf_compartment_2": {"status": "consistent", "evidence": "cam_high AFTER shows the
  middle shelf compartment on the desktop; no contradictory state or placement is observed."},
    "lipstick": {"status": "unobservable", "evidence": "No complete assembled lipstick object
  is clearly visible in AFTER views; only the lipstick tube/body on the desk can be identified
  ."},
    "lipstick_cap": {"status": "unobservable", "evidence": "AFTER views do not clearly show a
  separate lipstick cap; drawer contents relevant to a cap are not sufficiently visible."},
    "lipstick_tube": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
  show the lipstick tube/body standing on the desktop near the shelf/drawer area, matching ON
  desktop."},
    "waste_paper": {"status": "consistent", "evidence": "cam_high AFTER shows crumpled white
  paper inside the white desktop trash bin, matching INSIDE desktop_trash_bin."},
    "drawer_cabinet": {"status": "consistent", "evidence": "cam_high AFTER shows the white
  drawer cabinet on the desktop; static furniture is present, with layer states separately
  audited."},
    "drawer_1": {"status": "consistent", "evidence": "cam_high AFTER identifies the top drawer
   layer as closed/flush, agreeing with belief open=false."},
    "drawer_3": {"status": "unobservable", "evidence": "cam_high AFTER does not provide an
  unobstructed view of the lowest drawer front because the open second drawer and angle
  obscure it; state cannot be checked."},
    "drawer_2": {"status": "consistent", "evidence": "cam_high AFTER clearly shows the middle
  drawer tray pulled out, agreeing with belief open=true."},
    "desktop": {"status": "consistent", "evidence": "All AFTER cameras show the white desktop
  surface supporting the visible shelf, drawer cabinet, trash bin, pencil box, eraser,
  lipstick body, and other desk objects."},
```

```
    "desktop_trash_bin": {"status": "consistent", "evidence": "cam_high AFTER shows the white
   trash bin with black rim resting on the desktop and containing paper, matching ON desktop as
    a container."},
    "eraser": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER show the
    small yellow eraser resting on the desktop near the lipstick body, matching ON desktop."},
    "red_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER clearly shows the
   white/red pen grasped between the gripper jaws; this matches the belief that red_pen is held
   ."},
    "blue_pen": {"status": "unobservable", "evidence": "AFTER views do not show a distinct
   blue pen; the third drawer where it may be hidden is not open/visible enough to inspect."},
    "pencil_case": {"status": "consistent", "evidence": "cam_high and cam_left_wrist AFTER
   show the purple pencil box open on the desktop, matching belief open=true and ON desktop."},
    "black_pen": {"status": "consistent", "evidence": "cam_left_wrist AFTER shows a black/
   white pen lying inside the open purple pencil box, matching INSIDE pencil_case."}
   },
   "reason": "AFTER views show the successful grasp of the red pen and provide a clear high-
   camera audit of the drawer cabinet: top drawer closed, second drawer open, third drawer not
   sufficiently visible. Most placements match belief; the main additional mismatch is that the
    second interchangeable book is visually in the blue shelf rather than on the desktop.",
   "observed_state_delta": [
    {"operation": "observe_held", "node_id": "red_pen", "confidence": 0.94, "evidence": "
   cam_left_wrist AFTER shows the red/white pen firmly between the gripper jaws."},
    {"operation": "relocate", "node_id": "book_2", "relation": "INSIDE", "target": "
   shelf_compartment_2", "confidence": 0.82, "evidence": "cam_high AFTER shows a second
   distinct upright book/green-backed item within the blue shelf area; using interchangeable
   book bookkeeping, assign it to book_2 rather than the believed desktop placement."}
  ]
}
```


F.15 STEP 18

Action: `putin(red_pen,pencil_case)`

Logged action ID: `13f39058116c43d48525653e6eafba92`

Head camera | Left wrist camera | Right wrist camera

Before execution



After execution



Action-plan JSON.


```
{
  "reason": "The red pen is currently held, and the pencil case is visible and open. The goal
   requires the red pen to be inside the pencil case, so place the held red pen there.",
  "action": {"name": "putin", "object": "red_pen", "target": "pencil_case", "node_ids": ["
   red_pen", "pencil_case"]}
}
```

Evaluator JSON.


```
{"status": "success", "reason": "The red pen is no longer held and is visibly placed inside
  the purple pencil case/tray after the action."}
```


Execution record (selected backend fields).


```
{
  "action_effect_committed": true,
  "observation_patch_committed": false,
  "graph_update_committed": true,
  "goal_satisfied": false,
  "planning_progress": {"current_cost": 11.0, "best_cost": 11.0, "stagnant_steps": 0, "
  max_stagnant_steps": 10, "last_cost_change_step": 18, "enabled": true, "cost_before": 12.0,
  "cost_after": 11.0, "cost_changed": true, "improved_historical_best": true},
  "termination_triggered": false,
  "reconciliation_status": "ignored",
  "validation_errors": []
}
```


Step 18: global visual reconciliation JSON. The following complete response refers to the before/after observations for this action.


```
{
  "drawer_observations": {
    "drawer_1": {"confidence": 0.95, "identity_evidence": "cam_high AFTER shows the white
  drawer cabinet at left; the top slot/front is directly under the cabinet top and above the
  middle drawer front.", "observed_state": "closed", "state_evidence": "cam_high AFTER shows
  the top drawer front seated flush with no tray protruding."},
    "drawer_3": {"confidence": 0.9, "identity_evidence": "cam_high AFTER shows the same
  cabinet with fixed top and two visible fronts; the third layer is the bottom slot beneath
  the second/front row, partially visible below.", "observed_state": "unobservable", "
  state_evidence": "The bottom layer/front is largely hidden by the open middle drawer and
  camera angle, so its flush/open displacement cannot be established."},
    "drawer_2": {"confidence": 0.95, "identity_evidence": "cam_high AFTER shows the middle
  slot/front below the closed top drawer and above the lower area; cam_left_wrist AFTER
  matches this protruding middle tray/front.", "observed_state": "open", "state_evidence": "
  cam_high AFTER and cam_left_wrist AFTER show the middle drawer tray pulled outward with its
  front handle displaced from the cabinet frame."}
  },
  "object_observations": {
    "book_1": {"evidence": "cam_high AFTER shows one book standing in/against the first shelf
  compartment area beside the blue book rack; using interchangeable book assignment, this
  agrees with one book inside shelf_compartment_1 and closed/not held.", "status": "consistent
  "},
    "book_2": {"evidence": "cam_high AFTER shows a second distinct book/booklike item
  associated with the blue book rack compartments; assigned to book_2, its placement is
  consistent with being inside shelf_compartment_2 and not held.", "status": "consistent"},
    "bookshelf": {"evidence": "cam_high AFTER shows the blue shelf/book rack stationary on the
   desktop; furniture placement is consistent with the belief, though the belief marks it not
  visible/reachable symbolically.", "status": "consistent"},
    "shelf_compartment_1": {"evidence": "cam_high AFTER shows the blue rack compartment on the
   desktop with a book in the left/first compartment area; container remains on the table.", "
  status": "consistent"},
    "shelf_compartment_3": {"evidence": "cam_high AFTER shows the blue rack standing on the
  desktop; the third/right compartment is visible and stationary on the table.", "status": "
  consistent"},
    "shelf_compartment_2": {"evidence": "cam_high AFTER shows the blue rack standing on the
  desktop with the middle compartment area visible; container placement on the table agrees.",
   "status": "consistent"},
    "lipstick": {"evidence": "No assembled lipstick object is directly identifiable in AFTER
  views; only the tube/body on the desk is visible, so this whole-object location cannot be
  checked.", "status": "unobservable"},
    "lipstick_cap": {"evidence": "AFTER views do not clearly show the lipstick cap; the top
  drawer that would occlude it is closed, and no exact cap placement is established.", "status
  ": "unobservable"},
    "lipstick_tube": {"evidence": "cam_high AFTER and cam_left_wrist AFTER show the gold
  lipstick tube/body upright on the desktop near the front-center/right, matching ON desktop
  and not held.", "status": "consistent"},
    "waste_paper": {"evidence": "cam_high AFTER shows white crumpled paper inside the small
  trash bin on the desktop, matching INSIDE desktop_trash_bin.", "status": "consistent"},
    "drawer_cabinet": {"evidence": "cam_high AFTER shows the white drawer cabinet stationary
  on the desk; individual layer states are reported separately.", "status": "consistent"},
    "drawer_1": {"evidence": "cam_high AFTER identifies the top drawer as the upper slot under
   the cabinet top and shows its front flush closed, agreeing with belief open=false.", "
  status": "consistent"},
    "drawer_3": {"evidence": "cam_high AFTER shows the lower region only partly; the bottom
  drawer front is obscured by the protruding second drawer/angle, so its open/closed state and
   visibility cannot be verified.", "status": "unobservable"},
```

```
    "drawer_2": {"evidence": "cam_high AFTER and cam_left_wrist AFTER show the second/middle
   drawer pulled open, agreeing with belief open=true.", "status": "consistent"},
    "desktop": {"evidence": "All AFTER views show the white tabletop supporting the pencil
   case, drawer cabinet, book rack, trash bin, eraser, and lipstick tube; no contradiction to
   the surface belief.", "status": "consistent"},
    "desktop_trash_bin": {"evidence": "cam_high AFTER shows the small white trash bin with
   black rim on the desktop at the back-right, containing paper, matching ON desktop.", "status
   ": "consistent"},
    "eraser": {"evidence": "cam_high AFTER and cam_left_wrist AFTER show the small beige
   eraser resting directly on the desktop near center-front, matching ON desktop.", "status": "
   consistent"},
    "red_pen": {"evidence": "cam_high AFTER and cam_left_wrist AFTER clearly show the red pen
   lying inside the open purple pencil case, matching INSIDE pencil_case and not held.", "
   status": "consistent"},
    "blue_pen": {"evidence": "No blue pen is clearly visible in AFTER views; the believed
   hidden/no-location state cannot be visually confirmed.", "status": "unobservable"},
    "pencil_case": {"evidence": "cam_high AFTER and cam_left_wrist AFTER show the purple
   pencil case open on the desktop with pens inside, matching ON desktop and open=true.", "
   status": "consistent"},
    "black_pen": {"evidence": "cam_high AFTER and cam_left_wrist AFTER show a black/dark pen
   lying inside the open purple pencil case, matching INSIDE pencil_case.", "status": "
   consistent"}
   },
   "observed_state_delta": [],
   "reason": "AFTER images confirm the requested red pen placement inside the open pencil case
   and show no additional unambiguous object-location mismatches. The drawer audit identifies
   the fixed top layer closed, second/middle layer open, and third/bottom layer not
   sufficiently observable; no extra delta is needed because these agree with the current
   belief where observable."
}
```