SIMWORLD STUDIO: Automatic Environment Generation with Evolving Coding Agent for Embodied Agent Learning

Haoqiang Kang[1]* **Xiaokang Ye**[1]* **Yuhan Liu**[2] **Siddhant Hitesh Mantri**[1]
Lingjun Mao[1] **James Fleming**[1] **Drishti Regmi**[1] **Lianhui Qin**[1]
[1]UC San Diego [2]New York University

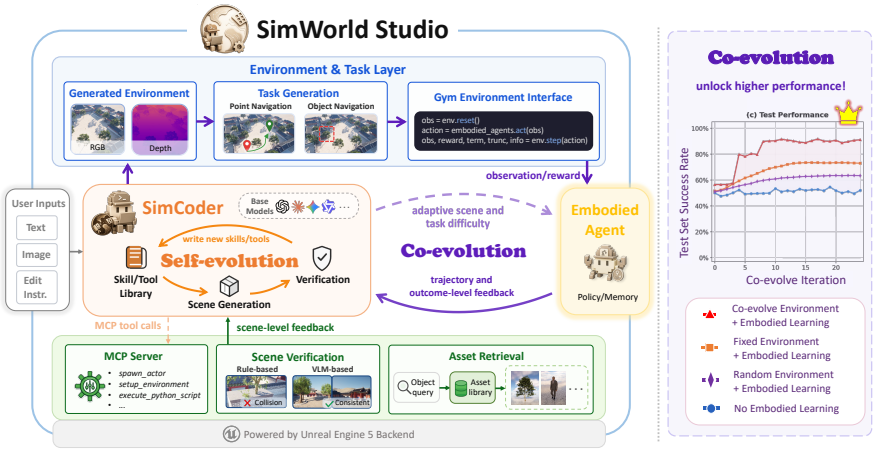


Figure 1: **SIMWORLD STUDIO: (Left)** SIMCODER automatically generates UE5 interactive environments with realistic 3D scenes, learning tasks, and Gym interfaces. **(Right)** Co-evolving environment generation with embodied learning substantially improves test success over both fixed-environment training and the untrained-agent baseline.

Abstract

Training embodied agents requires diverse, interactive environments with verifiable tasks and learning interfaces. We introduce SIMWORLD STUDIO, a platform built on Unreal Engine 5 that connects environment generation with embodied-agent learning. Its coding agent, SIMCODER, constructs scenes from language, images, or editing instructions through engine-level tools. Rule-based and vision-language verifiers guide repair, while successful procedures are stored as reusable tools and skills. Generated scenes are paired with tasks and exported through a Gymnasium-compatible interface. The platform also supports an adaptive curriculum: training-time agent performance controls environment difficulty, while failure trajectories inform the agent's in-context decision rules. Navigation experiments show improvements from verification and skill reuse, transfer to independently authored simulation environments, and an 18-percentage-point success-rate gain from co-evolution over fixed-difficulty learning. Additional evaluations cover 200 scenes per builder, cross-model judge agreement, generation cost, and a MuJoCo manipulation adapter. On 50 held-out pick-and-place scenes, experience distilled from generated tasks improves success from 40% to 50%. These results support adaptive environment generation for simulated embodied learning; real-robot transfer and larger-scale deployment remain open questions.

*Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

1 Introduction

Large language and vision models increasingly act in software environments: they write code, operate interfaces, and learn from interactive sandboxes [16, 26, 99]. Embodied agents likewise need environments in which they can perceive, act, and receive feedback [18, 104]. Constructing these environments is more demanding than producing plausible images or static geometry: scenes must support physical execution, reachable goals, task-dependent observations and actions, and verifiable outcomes.

Existing simulators provide essential infrastructure for navigation, manipulation, and driving [38, 54, 17, 40]. Procedural systems such as ProcTHOR [14] and Infinigen [59] scale scene creation through explicit generation rules. Language-conditioned systems, including Holodeck [87], also produce interactive environments. Recent coding agents further expand scene construction through executable programs [33, 98, 81, 53]. Our focus is integrating code-based construction, verification, reusable skills, task compilation, and learner feedback into one environment-generation workflow.

We introduce SIMWORLD STUDIO, built on the UE5-based SimWorld runtime [89]. Its coding agent, SIMCODER, constructs scenes from natural-language prompts, reference images, and editing requests. Structured tools expose engine operations and asset retrieval; geometric checks and multi-view VLM feedback guide iterative repair. When a corrective procedure is reusable, SIMCODER records it as a tool or skill for later generations. This *self-evolution* mechanism builds on tool-making and skill-accumulation approaches [11, 70], and is evaluated through component ablations.

The generated scenes become learning environments through automatic task construction and a Gymnasium-compatible interface. For navigation, the system samples start–goal pairs, checks NavMesh connectivity, and exposes observations, actions, and rewards. A MuJoCo adapter additionally instantiates tabletop pick-and-place tasks, providing an initial test of the same generate–verify–repair workflow on another backend and task family.

SIMWORLD STUDIO also closes a feedback loop between the environment generator and the embodied learner. In our navigation instantiation, aggregate success on generated training episodes gates an eight-level curriculum, while failed trajectories supply corrective rules to the frozen embodied model. This is adaptive curriculum learning over a specified difficulty space, rather than an unrestricted search over all possible tasks. The evaluation benchmark supplies no adaptation feedback.

Three navigation case studies assess scene quality, learning from generated environments, and adaptive co-evolution (Figure 3). Co-evolution reaches 90% success on SimWorld-MMNav, compared with 72% for fixed-difficulty learning and 50% for the unadapted baseline in that experiment. Additional experiments evaluate generation at 200 scenes per builder, cross-judge consistency, cost, and the separate effects of curriculum, scene resampling, and rule memory. A 50-training/50-test manipulation study provides preliminary evidence beyond navigation. We distinguish these measured results from throughput projections and from future real-world validation. Implementation details, interfaces, and examples appear in Appendices A, G, and H.

2 SIMWORLD STUDIO

SIMWORLD STUDIO is built on the open-source Unreal Engine 5 based SIMWORLD library [89] by inheriting its assets, runtime, and Python wrapper on the UE5 backend, UE5 supplies rendering, physics, and NavMesh; SimWorld supplies the embodied runtime, assets, and Python execution wrapper. The 17,955-asset registry is inherited infrastructure. SIMWORLD STUDIO adds the coding agent, evolving tools/skills, verification and repair, task compilation, and learner feedback. SIMWORLD STUDIO makes two main methodological contributions: **(1) Automatic environment generation** (§2.1): a coding agent that synthesizes executable 3D scenes, evolves its own skill and tool library from verifier feedback, and exports each scene as a Gymnasium-compatible embodied environment. **(2) Co-evolution as an adaptive curriculum mechanism** (§2.2): embodied agent performance is fed back into environment generation, so new environments target the agent's current weaknesses and remain near the boundary of its capabilities.

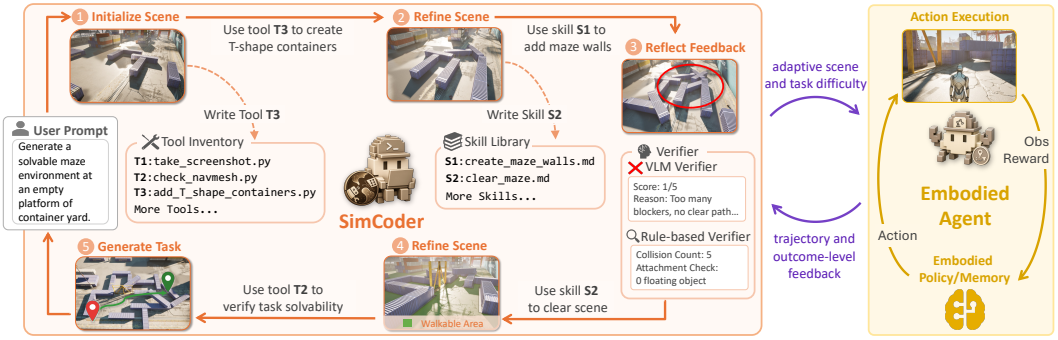


Figure 2: SIMCODER turns a user prompt into an interactive environment through an automatic self-evolving loop: it writes tools, creates reusable skills, reuses them across iterations, and refines the scene with verifier feedback. NavMesh-based tools are used to generate solvable navigation tasks. SIMCODER further uses embodied-agent feedback to autonomously adapt environment difficulty and co-evolve with the embodied agent (purple loop).

2.1 SIMCODER: Coding Agent for Automatic Environment Generation

As shown in Figure 1(Left), SIMWORLD STUDIO comprises three components: **SIMCODER**, an LLM coding agent that drives generation; *tool and skill libraries*, including an inventory of Python functions as tools and a library of skills which are reusable procedures, exposed through a Model Context Protocol [3] (MCP) bridge; and *verifiers* that return verification signals (rule- and VLM-based) to guide scene construction and revision. As illustrated in Figure 2 with a maze-generation task, generation flows in a loop: given a user prompt (text, image, or edit instruction), SIMCODER issues tool calls or skill retrievals through MCP; the backend executes them and returns a state update or a verifier signal, which SIMCODER consumes as the next observation and either continues building the scene, revises in place, or, when a fix proves broadly useful, writes a new tool or skill back into its library so future generations can reuse it. Once the scene passes verification, SIMCODER derives a task from it and exports it as a Gymnasium environment, allowing embodied agents to interact with it.

MCP Tools. Tools are Python function calls that SIMCODER invokes through the MCP bridge to act on the UE backend. The inventory has two parts. *Primitive tools* are the fixed, predefined set of operations needed to author a scene end-to-end (e.g., actor management, environment and asset management, and scene evaluation). *Extensible tools* cover everything outside this fixed set: a Python escape hatch runs arbitrary Unreal Engine Python for one-off operations, and any pattern that proves useful across runs is promoted via self-evolution into a named wrapper that the bridge registers as a first-class MCP tool, indistinguishable from the primitives at call time. Step 1 of Figure 2 shows one such wrapper (`add_T_shape_containers.py`) being invoked from the Tool Inventory. The full primitive inventory is in Appendix A.1.

Skill Library. Skills sit one layer above tools. Each skill is a Markdown document that records how to use a tool (or a sequence of tools) to accomplish a particular composition goal; SIMCODER retrieves applicable skills at the start of each episode and issues the underlying tool calls itself, so skills tell it how to compose tools rather than bypass them. As with tools, the library has two parts: a small set of *primitive skills* ships with the platform (covering common composition goals such as building placement, city layout, and screenshot capture for the VLM judge), and *extensible skills* accumulate over time through self-evolution. Step 2 of Figure 2 shows SIMCODER retrieving an evolved skill (`create_maze_walls.md`) to add walls to the partially-built maze.

Verification Loop. SIMWORLD STUDIO verifies generated scenes through two complementary verifiers (Step 3 of Figure 2). A *rule-based verifier* computes physical and geometric metrics (e.g., collisions, vertical support, in-bounds placement) from the scene graph and is invoked on every actor-modifying tool call. A *VLM-based verifier* captures multi-view screenshots and asks a vision-language model to score semantic alignment against the prompt, returning structured feedback after each block of construction. Verifier responses re-enter the trajectory as the next observation, and SIMCODER revises in place. In the maze episode of Figure 2, for example, the rule-based verifier

reports a collision count of 5 and the VLM verifier scores the scene 1/5 (*"too many blockers, no clear path..."*); SIMCODER then retrieves the `clear_maze.md` skill and removes the redundant containers before continuing. Full metric definitions are in Appendix C.2.

Self-evolution. Self-evolution turns one-off fixes into permanent capabilities. When a verifier failure recurs across attempts, SIMCODER restates the failure at the level of a *class* of cases and authors a new tool or skill that addresses the class rather than the specific instance, writing it to the registry so all subsequent runs can retrieve it [11, 56]. Step 4 of Figure 2 illustrates one such update: after the maze fails verification, SIMCODER writes a new skill (`clear_maze.md`) that generalizes the corrective procedure (i.e., removing redundant blockers from any container layout) and the skill is then available for all future episodes. Representative authored entries are in Appendix A.3.

Task Generation. SIMCODER also generates a task on top of a generated scene, using the same tool-call interface to query scene structure (e.g., NavMesh for traversable regions). Step 5 of Figure 2 shows the maze scene compiled into a navigation task with a sampled start–goal pair on the walkable area. We instantiate two canonical navigation families as a representative case: *point navigation* [2] (goal = coordinate) and *object navigation* [8] (goal = semantic target). Geometric reachability is checked using NavMesh connectivity, and verifiability follows from the same scene-query tools: during execution, we directly query the agent pose and target location and check success based on distance to the target.

Gymnasium Compilation. A generated environment then exports as a standard Gymnasium environment, with `env.reset()` and `env.step(action)` returning RGB-D observations, agent pose, and reward (top of Figure 1(Left)). Because the contract is the standard one, compatible RL algorithms (e.g., PPO [63]) or training-free LLM policy (e.g., ReAct [88]) can interact through this interface, making each generated scene a first-class training substrate for embodied agent learning.

2.2 Co-Evolution: An Adaptive Curriculum Mechanism

So far the generator runs open-loop: it produces environments without knowing how the embodied agent fares in them. Co-evolution closes this loop and turns environment generation from a one-shot content-creation problem into an adaptive curriculum mechanism, where the scenes generated for training evolve together with the agents learning inside them. One round alternates two updates: the embodied agent trains on a batch of SIMCODER-generated environments, and SIMCODER then updates based on the resulting performance before producing the next batch. The two agents update individually, through different mechanisms.

Embodied Agent Evolving. From the embodied agent's perspective, co-evolution differs from fixed-environment training only in that the scene distribution drifts between rounds; the agent's update rule is unchanged. SIMWORLD STUDIO reuses the Gym interface of §2.1 without modification, so an RL policy (e.g., PPO [63]) updates via standard policy gradients on the reward returned by `step()`, while an LLM-based policy updates through in-context mechanisms such as incremental rule accumulation or reflection-style memory [70, 65].

Coding Agent Evolving. SIMCODER's adaptation is in-context; its underlying LLM weights remain fixed. The platform can expose scene-level verifier feedback, outcome-level success statistics, and trajectory-level experience. Our evaluated recipe uses these signals differently: scene feedback repairs geometry and prompt alignment; aggregate training success controls the environment difficulty level; and individual trajectory failures update the embodied agent's decision rules. Thus, the measured curriculum does not directly map each failure type into a new scene structure. Engine-specific actor creation, rendering, physics queries, and execution are isolated behind backend tools; the MuJoCo experiment in §3.4 tests this boundary for one manipulation family.

Section 3.3 instantiates this recipe for navigation tasks, using outcome-level agent outcomes to adapt SIMCODER's difficulty schedule while the agent improves through incremental rule accumulation. The resulting adaptive curriculum outperforms fixed-environment training.



Figure 3: **Three case studies evaluating SIMWORLD STUDIO.** Case 1 evaluates SIMCODER's scene generation quality across settings and LLM backbones. Case 2 trains embodied navigation agents in generated environments. Case 3 studies co-evolution where SIMCODER and the embodied agent iteratively improve each other.

3 Experiments and Analysis

We analyze SIMWORLD STUDIO through three case studies of increasing scope (Figure 3): environment generation quality (§3.1), embodied agent learning in generated environments (§3.2), and co-evolution between the environment generation and the embodied agent (§3.3).

3.1 Case Study 1: Can SIMCODER generate valid and diverse environments?

This case study evaluates whether SIMCODER can generate diverse, physically plausible 3D environments from natural language prompts, reference images, and editing instructions. As illustrated in Figure 3 (case study 1 left), SIMCODER receives a text prompt (e.g., "build a residential neighborhood with parallel streets and a park"), invokes MCP tools to spawn and arrange assets in the UE5 environment, and iteratively refines the scene through screenshot-based verification (§2.1).

Settings. We evaluate across three settings of increasing complexity: **(S1)** *Text-to-Scene*: generate a scene from a natural language prompt alone; **(S2)** *Image+Text-to-Scene*: generate with an additional reference image (hand-drawn sketch or aerial photo); **(S3)** *Scene Editing*: modify an existing scene by adding, removing, or rearranging objects without rebuilding from scratch. Each setting is tested at three difficulty levels (easy, medium, hard), yielding 9 evaluation scenes total. We use the two-axis

Table 1: **Scene generation quality (averaged across difficulty levels).** All metrics $\in [0, 1]$ (↑). Individual metrics are reported without a cross-metric aggregate. Bold = best per column. Metric colors: quantity, physical validity, semantic, aesthetic.

	S1: Text-to-Scene						S2: Image+Text-to-Scene						S3: Scene Editing					
	Rule-Based				*VLM*		*Rule-Based*				*VLM*		*Rule-Based*				*VLM*	
LLM	*Count*	*Diversity*	*No Collision*	*Gravity*	*Fidelity*	*Aesthetics*	*Count*	*No Collision*	*Gravity*	*In-Bounds*	*Fidelity*	*Style*	*Preserve*	*Edit Count*	*No Collision*	*Coherence*	*Edit Compl.*	*Layout*
Qwen3.5-9B	.00	.50	.37	.58	.20	.13	.60	.33	.53	1.0	.13	.13	1.0	.00	.37	.23	.23	.20
Qwen3.5-27B	.17	.46	.99	1.0	.43	.37	.73	.75	.83	1.0	.38	.30	1.0	.00	.99	.47	.13	.50
Sonnet 4.6	.33	**.80**	**1.0**	1.0	.53	.37	.80	.81	.87	1.0	.50	.37	1.0	1.0	.98	.40	**.67**	.40
Opus 4.6	**.75**	.72	.99	1.0	**.63**	**.47**	**.83**	**.86**	**.92**	1.0	**.62**	**.47**	1.0	1.0	.98	**.50**	.50	**.53**

evaluation from §2.1: rule-based metrics for physical validity (e.g., collision-free placement, gravity consistency, in-bounds placement) and VLM-as-judge metrics for semantic alignment (e.g., prompt fidelity, spatial fidelity, layout aesthetics); full definitions are in Appendix C.2.

Base Models. We benchmark four LLM backbones, including Claude Opus 4.6 [5], Claude Sonnet 4.6 [6], and Qwen3.5-27B/9B [58], all through the Claude Code agent framework [4] with the same MCP tool interface, verification loop, and skill library (§2.1). All agents differ only in the underlying LLM, isolating the contribution of model capability from platform infrastructure.

3.1.1 Results

Table 1 reports performance averaged across difficulty levels; full breakdowns are in Appendix C.1.

Physical and semantic quality vary with model and setting. Opus 4.6 and Sonnet 4.6 achieve collision-free scores of at least 0.98 in S1 and S3, but lower scores in image-guided S2 (0.86 and 0.81). Opus 4.6 has the highest prompt fidelity in S1 (0.63) and S2 (0.62); stronger physical scores do not imply perfect semantic alignment. We report individual metrics rather than combining heterogeneous physical and semantic measures.

Ablation studies on key platform components. Figure 4 ablates three platform components beyond the vanilla coding agent: MCP tools, verification loop, and self-evolution. The evaluation is conducted on a held-out test set of 9 scenes across S1/S2/S3. First, we observe that the *vanilla coding agent* fails to construct reliable environments (scoring 0.16). Then, adding *customized MCP tools* raises quality to 0.45 (+0.29), providing the structured action space needed for reliable asset interaction. Moreover, adding the *verification loop* improves quality by +0.10, as iterative screenshot-based correction catches spatial errors that single-pass generation misses. We find that *self-evolution* can break the plateau shared by all static configurations, further raising a +0.21 quality improvement by accumulating reusable placement strategies across generations. These results support the contribution of all three components within the evaluated setting. Full ablation details are in Appendix C.3.

Larger evaluation, judge agreement, and cost. An additional 200 scenes per builder yield mean PF/SRF/LAES scores of 0.499 ± 0.031 for Qwen3.5-27B and 0.118 ± 0.021 for Qwen3.5-2B (SD across scenes). Three independent judge families give pairwise Pearson correlations of 0.821–0.830 and Spearman correlations of 0.805–0.836 on per-scene mean scores. The correlations support consistency of scene rankings across judge families. Two complete town-generation traces require 5.5/8.9 minutes and $3.94/$4.77 in model-API charges, excluding local compute. Full results, the generator comparison, and stage-level cost accounting appear in Appendix F.

3.2 Case Study 2: Can embodied agents learn useful navigation abilities from SIMWORLD STUDIO-generated environments?

We evaluate whether LLM-based embodied agents can learn navigation policies in SIMWORLD STUDIO-generated environments and generalize to unseen scenes. As shown in Figure 3 (Case Study 2), an agent navigates a SIMCODER-generated city from RGB, bearing, and distance observations.

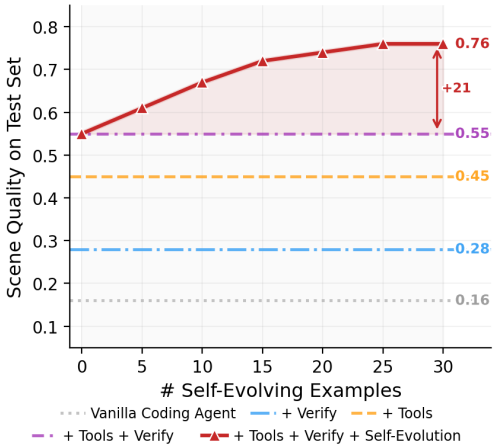


Figure 4: Ablation study results for SIMCODER in Case Study 1.

Table 2: **Embodied navigation on generated environments.** We report SR, SPL, SoftSPL, and nDTW. Metric colors denote task success, path efficiency, and trajectory fidelity. **Bold** marks the best result within each column; green marks improvement >0.05 over baseline.

LLM	Setting	Point Navigation				Object Navigation			
		SR(%, ↑)	SPL(↑)	SoftSPL(↑)	nDTW(↑)	SR(%, ↑)	SPL(↑)	SoftSPL(↑)	nDTW(↑)
Qwen3.5-27B	Baseline	20.21	0.0931	0.0882	0.2322	22.56	0.0795	0.0782	0.2486
	+ Memory	**26.76**	**0.1266**	**0.1036**	**0.2676**	**31.59**	**0.1429**	**0.1161**	**0.3159**
Qwen3.5-9B	Baseline	0.40	0.0039	0.1137	0.1137	0.00	0.0000	0.1160	0.1160
	+ Memory	**7.43**	**0.0722**	**0.2617**	**0.2617**	**14.62**	**0.1408**	**0.3059**	**0.3059**
Qwen3.5-2B	Baseline	1.38	0.0138	0.1030	**0.0034**	0.24	0.0024	0.1080	**0.0041**
	+ Memory	**6.23**	**0.0528**	**0.2119**	0.0022	**4.58**	**0.0389**	**0.1841**	0.0036

Navigation serves as a functional probe of environment quality, since failures may reflect either weak policies or defects in the generated world.

Settings. We consider two outdoor tasks: *Point Navigation* (PointNav), where the agent navigates to a target coordinate, and *Object Navigation* (ObjectNav), where it locates a target object category. Episodes are generated on SIMWORLD STUDIO maps via UE5 NavMesh by sampling start–goal pairs, computing shortest-path references, and filtering by reachability and path length. For each task, we sampled 1.2K episodes as our training set and evaluated agents on 329 held-out SIMWORLD STUDIO episodes; we additionally evaluate on the external SimWorld-MMNav benchmark [102] for cross-benchmark transfer. To assess how scene diversity affects downstream learning, we vary the number of training environments from 1 to 30 with the training budget fixed at 200 episodes.

Metrics. We report Success Rate (SR) for *task success*, Success weighted by Path Length (SPL) and Soft Success weighted by Path Length (SoftSPL) for *path efficiency*, and normalized Dynamic Time Warping (nDTW) for *trajectory fidelity*. Definitions are in Appendix D.3.

Models. We evaluate Qwen3.5-2B/9B/27B [58] with and without agent memory. **Memory keeps model weights frozen**: it distills trajectory experience into rules retrieved in context, rather than performing supervised fine-tuning. Our memory design is inspired by Generative Agents [50] and ExpeL [99], but extends them with *multi-level* updates at step, trajectory, and task granularities, allowing distilled strategies to be retrieved at the appropriate scope during inference (Appendix D.4).

3.2.1 Results

SIMWORLD STUDIO improves the embodied agents in both held-out environments and external benchmark. On held-out SIMWORLD STUDIO test environments (Table 2), agents trained in SIMWORLD STUDIO substantially outperform the no-training baseline across all model scales (e.g., Qwen3.5-9B: +14.62 percentage points on ObjectNav). These gains transfer to the external SimWorld-MMNav benchmark (Figure 5(Right); +12.0pp on 2B), providing evidence of simulation-to-simulation transfer.

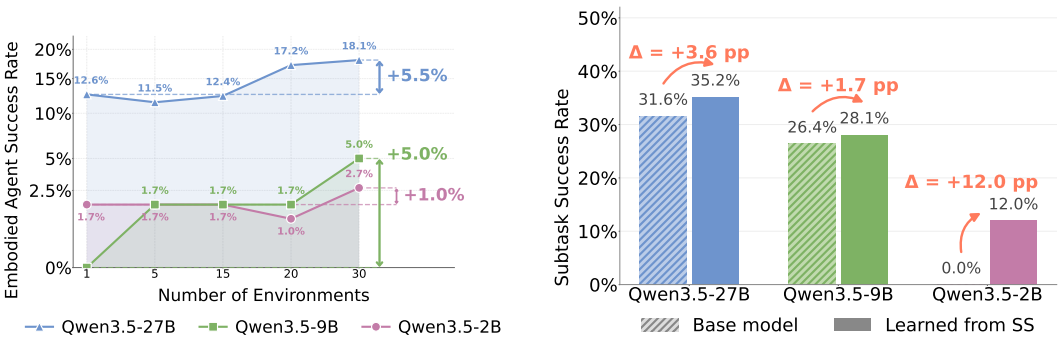


Figure 5: **Generalization analysis. (Left)** More diverse SIMWORLD STUDIO environments yield stronger test-time generalization. **(Right)** Embodied agents learned in SIMWORLD STUDIO transfer to SimWorld-MMNav across model scales.

SIMWORLD STUDIO's open-ended environment generation capability translates into stronger embodied agent generalization. With the training budget held fixed, Qwen3.5-27B test success improves as the number of distinct training environments increases (Figure 5(Left); $+5.5$pp on Qwen3.5-27B). The effect varies across model sizes. Observation-modality results likewise depend on model and split; RGB-D does not uniformly outperform RGB-only or depth-only inputs (Appendix D.5). A separate Qwen3.5-9B trajectory-SFT control reaches 6.80%/12.90% PointNav/ObjectNav SR, compared with 7.43%/14.62% for memory and 0.40%/0.00% without adaptation (Appendix F.5).

3.3 Case Study 3: Can SIMCODER co-evolve with embodied agents with agent feedback?

Case Studies 1 and 2 evaluate SIMCODER and the embodied agent in isolation. As a demonstration of the platform's closed-loop capability, Case Study 3 instantiates a simple performance-gated adaptive curriculum. As shown in Figure 6 (a), SIMCODER generates progressively harder environments while the embodied agent updates its policy from experience, with each agent's output steering the other's next step via the co-evolution framework of §2.2.

Method. We instantiate a specific adaptive curriculum under the co-evolution framework of §2.2. Navigation difficulty is parameterized using path length, heading offset, and obstacle density, which are jointly quantized into eight levels of increasing challenge. Each co-evolution round proceeds in three explicit steps: **(i) Environment Generation:** SIMCODER generates a batch of navigation episodes at the current difficulty level. **(ii) Embodied Agent Learning:** The embodied agent attempts the batch and records any failed episodes. It then distills these failure trajectories into a small set of prioritized decision rules (e.g., *"if the goal lies behind the agent, turn around before exploring forward"*). Crucially, it *appends* these to its existing rule set. This incremental update retains earlier strategies while adding failure-derived corrections [1]. **(iii) Environment Adaptation:** SIMCODER evaluates the agent's mean Success rate on the batch. If the agent clears a level-specific mastery threshold, SIMCODER advances the next batch of environments to a higher difficulty; otherwise, it holds the current level until the agent catches up. Full details are in Appendix E.

Setting. We evaluate on the external SimWorld-MMNav benchmark [102] and compare four conditions using Qwen3.5-9B: **(1)** co-evolving environment (our full closed-loop system), **(2)** fixed-difficulty environment (held constant at level 3), **(3)** random-difficulty environment (random level from 0 to 7), and **(4)** base model (no embodied agent learning).

3.3.1 Results

Adaptive curricula drive continuous improvement and prevent early saturation. The training dynamics of the co-evolving system (Figure 6b) exhibit a characteristic *drop-and-recover* pattern [69]: the agent's Success rate dips exactly when SIMCODER introduces harder tasks, but recovers rapidly as distilled strategies transfer to the new difficulty tier. Ultimately, this closed-loop co-evolution

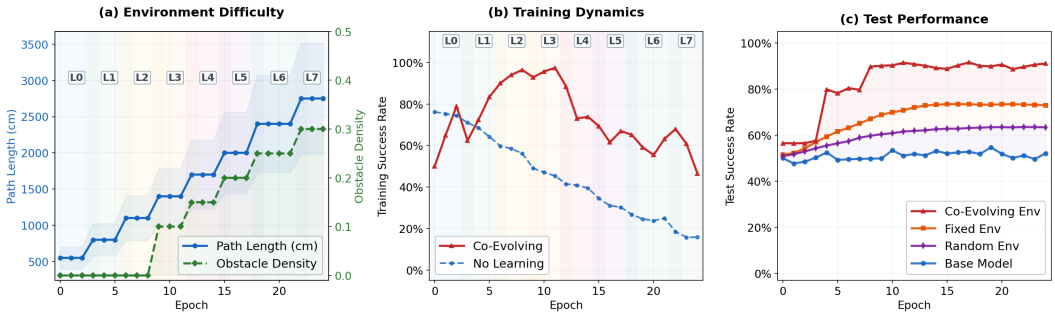


Figure 6: **Co-evolution of SIMCODER and embodied agent.** **(a)** Environment difficulty across 8 levels. **(b)** Training dynamics: the co-evolving agent drops at each level transition then recovers. **(c)** Test performance on the SimWorld-MMNav benchmark.

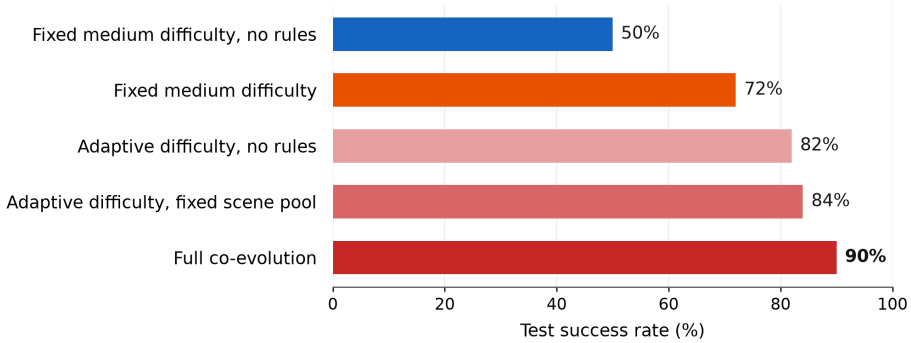


Figure 7: **Co-evolution factor ablations.** Qwen3.5-9B on SimWorld-MMNav. All settings use rule memory and scene resampling unless indicated otherwise. Fixed medium difficulty is L3 (10–18 m paths, 0–60° heading offset, obstacle density 0.15); adaptive difficulty advances after mastery.

yields a 90% test Success rate on SimWorld-MMNav. This is an 18-percentage-point gain over the fixed-environment learning baseline (72%), which plateaus in the observed run, and a 40-point gain over the untrained baseline (50%; Figure 6c).

Separating curriculum, diversity, and rule memory. Figure 7 holds the training-episode budget and the 329-episode external benchmark fixed. With resampling and rules enabled, adaptive difficulty improves SR from 72% to 90%; removing rules from the adaptive condition reduces SR to 82%, and retaining a fixed scene pool reduces it to 84%. These contrasts separate the evaluated effects of difficulty scheduling, scene resampling, and rule memory. Curriculum gates use only generated training outcomes; no test trajectory or outcome is used for adaptation. Failure-specific rules and an example level transition are detailed in Appendix F.6.

3.4 Beyond Navigation: MuJoCo Manipulation

A backend adapter retains task generation, verification/repair, and Gym evaluation while replacing UE5 execution with MuJoCo. SIMCODER generates tabletop pick-and-place environments with 0, 1, or 2 obstacles, including object placement, robot-control interfaces, and success predicates. Checks cover reachability, transport distance, gripper clearance, and path blocking. A scripted expert and a Qwen3.5-27B policy using images and object poses solve all 100 generated tasks in a feasibility test.

A separate learning study uses 50 training scenes and 50 disjoint test scenes. Both Qwen3.5-27B conditions omit the hand-written strategy block; only the adapted condition uses experience distilled from generated training tasks. Success rises from 20/50 to 25/50, while final distance, steps, and collisions decrease (Table 3). The feasibility study tests task solvability, whereas the separate learning study tests experience reuse without a hand-written strategy block. This provides initial evidence for one manipulation family in simulation.

Table 3: **Held-out manipulation results.** Same 50 test scenes; weights remain frozen. Distance is median final placement distance in meters; steps and collisions are means.

Training experience	Success	Distance	Steps	Collisions
None	20/50 (40%)	0.084	8.2	87.7
Distilled from 50 scenes	25/50 (50%)	0.040	5.4	52.3

4 Related Work

Embodied simulation and scene generation. Interactive simulators cover indoor navigation, household interaction, driving, and urban robotics [38, 40, 54, 41, 17, 79]. ProcTHOR and Infinigen scale scene creation procedurally [14, 59]; Holodeck generates interactive indoor scenes from language [87]; and agentic systems construct scenes through code or structured tools [33, 81, 98, 53]. SIMWORLD STUDIO combines engine-level generation, reusable skill evolution, verification, task export, and adaptive curricula. Backend and task differences complicate direct generator comparisons (Appendix F.4).

Agent–environment adaptation. Unsupervised environment design and POET adapt training environments to agent capabilities [15, 35, 51, 72, 61]. LLM-based methods generate or revise rewards, tasks, and environments [26, 47, 19, 42, 92]. Agent-World uses MCP-backed digital environments for agent training [16]. Our contribution integrates these ideas with verified 3D scene construction and embodied interfaces, and evaluates a bounded mastery-gated navigation curriculum. Additional discussion and a qualitative platform comparison are in Appendix B.

5 Conclusion and Limitations

SIMWORLD STUDIO integrates coding-agent scene construction, verification, reusable skills, and task export with adaptive embodied learning. Navigation results and an initial MuJoCo manipulation study show that generated environments can provide useful experience. The evidence is simulation-to-simulation: no real-robot transfer is established. Generation still depends on model spatial reasoning, asset coverage, and verifier quality. Cross-VLM agreement does not replace human validation. Our generation evaluation covers 200 scenes per builder, and the manipulation study covers one tabletop task family. Our eight-level curriculum explores a predefined difficulty space; long-horizon tasks, multi-agent learning, and real-world deployment remain future work.

Acknowledgment

We sincerely acknowledge Aria Lin for her contributions to the related work investigation. We also thank Lingge Meng, Vinayak Sharma, and Ishan Vaish for their support in running exploratory experiments, recording demonstration videos and assisting with the development of the project.

References

[1] Lakshya A Agrawal, Shangyin Tan, Dilara Soylu, Noah Ziems, Rishi Khare, Krista Opsahl-Ong, Arnav Singhvi, Herumb Shandilya, Michael J Ryan, Meng Jiang, Christopher Potts, Koushik Sen, Alexandros G. Dimakis, Ion Stoica, Dan Klein, Matei Zaharia, and Omar Khattab. Gepa: Reflective prompt evolution can outperform reinforcement learning, 2026. URL https://arxiv.org/abs/2507.19457.

[2] Peter Anderson, Angel Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, and Amir R. Zamir. On evaluation of embodied navigation agents, 2018. URL https://arxiv.org/abs/1807.06757.

[3] Anthropic. Introducing the model context protocol. https://www.anthropic.com/news/model-context-protocol, 2024. Published November 25, 2024. Accessed: 2026-05-03.

[4] Anthropic. Claude code by anthropic. https://www.anthropic.com/product/claude-code, 2026. Accessed: 2026-05-07.

[5] Anthropic. Introducing claude opus 4.6. https://www.anthropic.com/news/claude-opus-4-6, February 2026. Accessed: 2026-05-07.

[6] Anthropic. Claude sonnet 4.6. https://www.anthropic.com/claude/sonnet, February 2026. Accessed: 2026-05-07.

[7] Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, et al. Program synthesis with large language models. *arXiv preprint arXiv:2108.07732*, 2021. URL https://arxiv.org/abs/2108.07732.

[8] Dhruv Batra, Aaron Gokaslan, Aniruddha Kembhavi, Oleksandr Maksymets, Roozbeh Mottaghi, Manolis Savva, Alexander Toshev, and Erik Wijmans. Objectnav revisited: On evaluation of embodied agents navigating to objects. *CoRR*, abs/2006.13171, 2020. URL https://arxiv.org/abs/2006.13171.

[9] Aleksey Bokhovkin, Quan Meng, Shubham Tulsiani, and Angela Dai. Scenefactor: Factored latent 3d diffusion for controllable 3d scene generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 628–639, 2025.

[10] Jake Bruce, Michael D Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, et al. Genie: Generative interactive environments. In *Forty-first International Conference on Machine Learning*, 2024.

[11] Tianle Cai, Xuezhi Wang, Tengyu Ma, Xinyun Chen, and Denny Zhou. Large language models as tool makers, 2024.

[12] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde De Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021. URL https://arxiv.org/abs/2107.03374.

[13] Jaeyoung Chung, Suyoung Lee, Hyeongjin Nam, Jaerin Lee, and Kyoung Mu Lee. Luciddreamer: Domain-free generation of 3d gaussian splatting scenes, 2023. URL https://arxiv.org/abs/2311.13384.

[14] Matt Deitke, Eli VanderBilt, Alvaro Herrasti, Luca Weihs, Jordi Salvador, Kiana Ehsani, Winson Han, Eric Kolve, Ali Farhadi, Aniruddha Kembhavi, and Roozbeh Mottaghi. Procthor: Large-scale embodied ai using procedural generation, 2022. URL https://arxiv.org/abs/2206.06994.

[15] Michael Dennis, Natasha Jaques, Eugene Vinitsky, Alexandre Bayen, Stuart Russell, Andrew Critch, and Sergey Levine. Emergent complexity and zero-shot transfer via unsupervised environment design. *Advances in neural information processing systems*, 33:13049–13061, 2020.

[16] Guanting Dong, Junting Lu, Junjie Huang, Wanjun Zhong, Longxiang Liu, Shijue Huang, Zhenyu Li, Yang Zhao, Xiaoshuai Song, Xiaoxi Li, Jiajie Jin, Yutao Zhu, Hanbin Wang, Fangyu Lei, Qinyu Luo, Mingyang Chen, Zehui Chen, Jiazhan Feng, Ji-Rong Wen, and Zhicheng Dou. Agent-world: Scaling real-world environment synthesis for evolving general agent intelligence, 2026. URL https://arxiv.org/abs/2604.18292.

[17] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator, 2017. URL https://arxiv.org/abs/1711.03938.

[18] Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. Palm-e: An embodied multimodal language model, 2023. URL https://arxiv.org/abs/2303.03378.

[19] Maxence Faldor, Jenny Zhang, Antoine Cully, and Jeff Clune. Omni-epic: Open-endedness via models of human notions of interestingness with environments programmed in code. *arXiv preprint arXiv:2405.15568*, 2024. URL https://arxiv.org/abs/2405.15568.

[20] Linxi Fan, Guanzhi Wang, Yunfan Jiang, Ajay Mandlekar, Yuncong Yang, Haoyi Zhu, Andrew Tang, De-An Huang, Yuke Zhu, and Anima Anandkumar. Minedojo: Building open-ended embodied agents with internet-scale knowledge, 2022. URL https://arxiv.org/abs/2206.08853.

[21] Jinyuan Fang, Yanwen Peng, Xi Zhang, Yingxu Wang, Xinhao Yi, Guibin Zhang, Yi Xu, Bin Wu, Siwei Liu, Zihao Li, et al. A comprehensive survey of self-evolving ai agents: A new paradigm bridging foundation models and lifelong agentic systems. *arXiv preprint arXiv:2508.07407*, 2025. URL https://arxiv.org/abs/2508.07407.

[22] Tianqing Fang, Hongming Zhang, Zhisong Zhang, Kaixin Ma, Wenhao Yu, Haitao Mi, and Dong Yu. Webevolver: Enhancing web agent self-improvement with co-evolving world model. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 8970–8986, 2025.

[23] Chuang Gan, Jeremy Schwartz, Seth Alter, Damian Mrowca, Martin Schrimpf, James Traer, Julian De Freitas, Jonas Kubilius, Abhishek Bhandwaldar, Nick Haber, Megumi Sano, Kuno Kim, Elias Wang, Michael Lingelbach, Aidan Curtis, Kevin Feigelis, Daniel M. Bear, Dan Gutfreund, David Cox, Antonio Torralba, James J. DiCarlo, Joshua B. Tenenbaum, Josh H. McDermott, and Daniel L. K. Yamins. Threedworld: A platform for interactive multi-modal physical simulation, 2021. URL https://arxiv.org/abs/2007.04954.

[24] Chen Gao, Baining Zhao, Weichen Zhang, Jun Zhang, Jinzhu Mao, Zhiheng Zheng, Fanhang Man, Jianjie Fang, Zile Zhou, Jinqiang Cui, Xinlei Chen, and Yong Li. Embodiedcity: A benchmark platform for embodied agent in real-world city environment. *arXiv preprint*, 2024.

[25] Huan-ang Gao, Jiayi Geng, Wenyue Hua, Mengkang Hu, Xinzhe Juan, Hongzhang Liu, Shilong Liu, Jiahao Qiu, Xuan Qi, Yiran Wu, et al. A survey of self-evolving agents: What, when, how, and where to evolve on the path to artificial super intelligence. *arXiv preprint arXiv:2507.21046*, 2025. URL https://arxiv.org/abs/2507.21046.

[26] Jiacheng Guo, Ling Yang, Peter Chen, Qixin Xiao, Yinjie Wang, Xinzhe Juan, Jiahao Qiu, Ke Shen, and Mengdi Wang. Genenv: Difficulty-aligned co-evolution between llm agents and environment simulators. *arXiv preprint arXiv:2512.19682*, 2025. URL https://arxiv.org/abs/2512.19682.

[27] Yanjiang Guo, Tony Lee, Lucy Xiaoyang Shi, Jianyu Chen, Percy Liang, and Chelsea Finn. Vlaw: Iterative co-improvement of vision-language-action policy and world model. *arXiv preprint arXiv:2602.12063*, 2026. URL https://arxiv.org/abs/2602.12063.

[28] Hongliang He, Wenlin Yao, Kaixin Ma, Wenhao Yu, Hongming Zhang, Tianqing Fang, Zhenzhong Lan, and Dong Yu. Openwebvoyager: Building multimodal web agents via iterative real-world exploration, feedback and optimization. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 27545–27564, 2025.

[29] Lukas Höllein, Ang Cao, Andrew Owens, Justin Johnson, and Matthias Nießner. Text2room: Extracting textured 3d meshes from 2d text-to-image models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7909–7920, 2023.

[30] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, et al. Metagpt: Meta programming for a multi-agent collaborative framework. In *The twelfth international conference on learning representations*, 2023.

[31] Yicong Hong, Yiqun Mei, Chongjian Ge, Yiran Xu, Yang Zhou, Sai Bi, Yannick Hold-Geoffroy, Mike Roberts, Matthew Fisher, Eli Shechtman, et al. Relic: Interactive video world model with long-horizon memory. *arXiv preprint arXiv:2512.04040*, 2025. URL https://arxiv.org/abs/2512.04040.

[32] Shengran Hu, Cong Lu, and Jeff Clune. Automated design of agentic systems. *arXiv preprint arXiv:2408.08435*, 2024. URL https://arxiv.org/abs/2408.08435.

[33] Ziniu Hu, Ahmet Iscen, Aashi Jain, Thomas Kipf, Yisong Yue, David A. Ross, Cordelia Schmid, and Alireza Fathi. Scenecraft: An llm agent for synthesizing 3d scene as blender code, 2024. URL https://arxiv.org/abs/2403.01248.

[34] Chengsong Huang, Wenhao Yu, Xiaoyang Wang, Hongming Zhang, Zongxia Li, Ruosen Li, Jiaxin Huang, Haitao Mi, and Dong Yu. R-zero: Self-evolving reasoning llm from zero data. 2025. URL https://arxiv.org/abs/2508.05004.

[35] Minqi Jiang, Edward Grefenstette, and Tim Rocktäschel. Prioritized Level Replay, 2020.

[36] Zhennan Jiang, Shangqing Zhou, Yutong Jiang, Zefang Huang, Mingjie Wei, Yuhui Chen, Tianxing Zhou, Zhen Guo, Hao Lin, Quanlu Zhang, Yu Wang, Haoran Li, Chao Yu, and Dongbin Zhao. Wovr: World models as reliable simulators for post-training vla policies with rl. *arXiv preprint arXiv:2602.13977*, 2026. URL https://arxiv.org/abs/2602.13977.

[37] Carlos E Jimenez, John Yang, Alexander Wettig, Shunyu Yao, Kexin Pei, Ofir Press, and Karthik Narasimhan. Swe-bench: Can language models resolve real-world github issues? *arXiv preprint arXiv:2310.06770*, 2023. URL https://arxiv.org/abs/2310.06770.

[38] Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Matt Deitke, Kiana Ehsani, Daniel Gordon, Yuke Zhu, Aniruddha Kembhavi, Abhinav Gupta, and Ali Farhadi. Ai2-thor: An interactive 3d environment for visual ai, 2022. URL https://arxiv.org/abs/1712.05474.

[39] Zhengfei Kuang, Rui Lin, Long Zhao, Gordon Wetzstein, Saining Xie, and Sanghyun Woo. Vulcan: Tool-augmented multi agents for iterative 3d object arrangement, 2026. URL https://arxiv.org/abs/2512.22351.

[40] Chengshu Li, Fei Xia, Roberto Martín-Martín, Michael Lingelbach, Sanjana Srivastava, Bokui Shen, Kent Vainio, Cem Gokmen, Gokul Dharan, Tanish Jain, et al. igibson 2.0: Object-centric simulation for robot learning of everyday household tasks. *arXiv preprint arXiv:2108.03272*, 2021. URL https://arxiv.org/abs/2108.03272.

[41] Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen Wang, Gabrael Levine, Wensi Ai, Benjamin Martinez, Hang Yin, Michael Lingelbach, Minjune Hwang, Ayano Hiranaka, Sujay Garlanka, Arman Aydin, Sharon Lee, Jiankai Sun, Mona Anvari, Manasi Sharma, Dhruva Bansal, Samuel Hunter, Kyu-Young Kim, Alan Lou, Caleb R Matthews, Ivan Villa-Renteria, Jerry Huayang Tang, Claire Tang, Fei Xia, Yunzhu Li, Silvio Savarese, Hyowon Gweon, C. Karen Liu, Jiajun Wu, and Li Fei-Fei. Behavior-1k: A human-centered, embodied ai benchmark with 1,000 everyday activities and realistic simulation, 2024. URL https://arxiv.org/abs/2403.09227.

[42] William Liang, Sam Wang, Hung-Ju Wang, Osbert Bastani, Dinesh Jayaraman, and Yecheng Jason Ma. Eurekaverse: Environment curriculum generation via large language models. *arXiv preprint arXiv:2411.01775*, 2024. URL https://arxiv.org/abs/2411.01775.

[43] Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, et al. Agentbench: Evaluating llms as agents. *arXiv preprint arXiv:2308.03688*, 2023. URL https://arxiv.org/abs/2308.03688.

[44] Xiaokang Liu, Zechen Bai, Hai Ci, Kevin Yuchen Ma, and Mike Zheng Shou. World-vla-loop: Closed-loop learning of video world model and vla policy. *arXiv preprint arXiv:2602.06508*, 2026. URL https://arxiv.org/abs/2602.06508.

[45] Zishan Liu, Zecong Tang, RuoCheng Wu, Xinzhe Zheng, Jingyu Hu, Ka-Hei Hui, Haoran Xie, Bo Dai, and Zhengzhe Liu. Imagine a city: Citygenagent for procedural 3d city generation, 2026. URL https://arxiv.org/abs/2602.05362.

[46] Calvin Luo, Zilai Zeng, Mingxi Jia, Yilun Du, and Chen Sun. Self-adapting improvement loops for robotic learning. *arXiv preprint arXiv:2506.06658*, 2025. URL https://arxiv.org/abs/2506.06658.

[47] Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Eureka: Human-level reward design via coding large language models. *arXiv preprint arXiv:2310.12931*, 2023. URL https://arxiv.org/abs/2310.12931.

[48] Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. Self-refine: Iterative refinement with self-feedback. *Advances in neural information processing systems*, 36:46534–46594, 2023.

[49] NVIDIA et al. Cosmos world foundation model platform for physical ai. *arXiv preprint arXiv:2501.03575*, 2025. URL https://arxiv.org/abs/2501.03575.

[50] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior, 2023. URL https://arxiv.org/abs/2304.03442.

[51] Jack Parker-Holder, Minqi Jiang, Michael Dennis, Mikayel Samvelyan, Jakob Foerster, Edward Grefenstette, and Tim Rocktäschel. Evolving curricula with regret-based environment design. In *International Conference on Machine Learning*, pages 17473–17498. PMLR, 2022.

[52] Despoina Paschalidou, Amlan Kar, Maria Shugrina, Karsten Kreis, Andreas Geiger, and Sanja Fidler. Atiss: Autoregressive transformers for indoor scene synthesis. *Advances in neural information processing systems*, 34:12013–12026, 2021.

[53] Nicholas Pfaff, Thomas Cohn, Sergey Zakharov, Rick Cory, and Russ Tedrake. Scenesmith: Agentic generation of simulation-ready indoor scenes. *arXiv preprint arXiv:2602.09153*, 2026. URL https://arxiv.org/abs/2602.09153.

[54] Xavier Puig, Eric Undersander, Andrew Szot, Mikael Dallaire Cote, Tsung-Yen Yang, Ruslan Partsey, Ruta Desai, Alexander William Clegg, Michal Hlavac, So Yeon Min, Vladimír Vondruš, Theophile Gervet, Vincent-Pierre Berges, John M. Turner, Oleksandr Maksymets, Zsolt Kira, Mrinal Kalakrishnan, Jitendra Malik, Devendra Singh Chaplot, Unnat Jain, Dhruv Batra, Akshara Rai, and Roozbeh Mottaghi. Habitat 3.0: A co-habitat for humans, avatars and robots, 2023. URL https://arxiv.org/abs/2310.13724.

[55] Zehan Qi, Xiao Liu, Iat Long Iong, Hanyu Lai, Xueqiao Sun, Wenyi Zhao, Yu Yang, Xinyue Yang, Jiadai Sun, Shuntian Yao, et al. Webrl: Training llm web agents via self-evolving online curriculum reinforcement learning. *arXiv preprint arXiv:2411.02337*, 2024. URL https://arxiv.org/abs/2411.02337.

[56] Cheng Qian, Chi Han, Yi Fung, Yujia Qin, Zhiyuan Liu, and Heng Ji. CREATOR: Tool creation for disentangling abstract and concrete reasoning of large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6922–6939. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.findings-emnlp.462. URL https://aclanthology.org/2023.findings-emnlp.462/.

[57] Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, Sihan Zhao, Lauren Hong, Runchu Tian, Ruobing Xie, Jie Zhou, Mark Gerstein, Dahai Li, Zhiyuan Liu, and Maosong Sun. Toolllm: Facilitating large language models to master 16000+ real-world apis, 2023. URL https://arxiv.org/abs/2307.16789.

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

[59] Alexander Raistrick, Lahav Lipson, Zeyu Ma, Lingjie Mei, Mingzhe Wang, Yiming Zuo, Karhan Kayan, Hongyu Wen, Beining Han, Yihan Wang, et al. Infinite photorealistic worlds using procedural generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12630–12641, 2023.

[60] Maxime Robeyns, Martin Szummer, and Laurence Aitchison. A self-improving coding agent. *arXiv preprint arXiv:2504.15228*, 2025. URL https://arxiv.org/abs/2504.15228.

[61] Mikayel Samvelyan, Akbir Khan, Michael Dennis, Minqi Jiang, Jack Parker-Holder, Jakob Foerster, Roberta Raileanu, and Tim Rocktäschel. Maestro: Open-ended environment design for multi-agent reinforcement learning. *arXiv preprint arXiv:2303.03376*, 2023. URL https://arxiv.org/abs/2303.03376.

[62] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools, 2023. URL https://arxiv.org/abs/2302.04761.

[63] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017. URL https://arxiv.org/abs/1707.06347.

[64] Ansh Kumar Sharma, Yixiang Sun, Ninghao Lu, Yunzhe Zhang, Jiarao Liu, and Sherry Yang. World-gymnast: Training robots with reinforcement learning in a world model, 2026. URL https://arxiv.org/abs/2602.02454.

[65] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, 36:8634–8652, 2023.

[66] Kabir Swain, Sijie Han, Ayush Raina, Jin Zhang, Shuang Li, Michael Stopa, and Antonio Torralba. Virtualenv: A platform for embodied ai research, 2026. URL https://arxiv.org/abs/2601.07553.

[67] Jiapeng Tang, Yinyu Nie, Lev Markhasin, Angela Dai, Justus Thies, and Matthias Nießner. Diffuscene: Denoising diffusion models for generative indoor scene synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20507–20518, 2024.

[68] GigaBrain Team, Boyuan Wang, Bohan Li, Chaojun Ni, et al. Gigabrain-0.5m*: a vla that learns from world model-based reinforcement learning. *arXiv preprint arXiv:2602.12099*, 2026. doi: 10.48550/arXiv.2602.12099. URL https://arxiv.org/abs/2602.12099.

[69] Lev S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA, 1978.

[70] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models, 2023.

[71] Hanqing Wang, Jiahe Chen, Wensi Huang, Qingwei Ben, Tai Wang, Boyu Mi, Tao Huang, Siheng Zhao, Yilun Chen, Sizhe Yang, Peizhou Cao, Wenye Yu, Zichao Ye, Jialun Li, Junfeng Long, Zirui Wang, Huiling Wang, Ying Zhao, Zhongying Tu, Yu Qiao, Dahua Lin, and Jiangmiao Pang. Grutopia: Dream general robots in a city at scale, 2024. URL https://arxiv.org/abs/2407.10943.

[72] Rui Wang, Joel Lehman, Jeff Clune, and Kenneth O. Stanley. POET: Open-Ended Coevolution of Environments and Their Optimized Solutions. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pages 142–151. ACM, 2019. doi: 10.1145/3321707.3321799.

[73] Rui Wang, Joel Lehman, Aditya Rawal, Jiale Zhi, Yulun Li, Jeff Clune, and Kenneth O Stanley. Enhanced poet: Open-ended reinforcement learning through unbounded invention of learning challenges and their solutions. In *Proceedings of the 37th International Conference on Machine Learning*, pages 9940–9951. PMLR, 2020. URL http://proceedings.mlr.press/v119/wang20l/wang20l.pdf.

[74] Xingyao Wang, Yangyi Chen, Lifan Yuan, Yizhe Zhang, Yunzhu Li, Hao Peng, and Heng Ji. Executable code actions elicit better llm agents. In *Forty-first International Conference on Machine Learning*, 2024.

[75] Xingyao Wang, Boxuan Li, Yufan Song, Frank F Xu, Xiangru Tang, Mingchen Zhuge, Jiayi Pan, Yueqi Song, Bowen Li, Jaskirat Singh, et al. Openhands: An open platform for ai software developers as generalist agents. *arXiv preprint arXiv:2407.16741*, 2024. URL https://arxiv.org/abs/2407.16741.

[76] Yufei Wang, Zhou Xian, Feng Chen, Tsun-Hsuan Wang, Yian Wang, Katerina Fragkiadaki, Zackory Erickson, David Held, and Chuang Gan. Robogen: Towards unleashing infinite data for automated robot learning via generative simulation. *arXiv preprint arXiv:2311.01455*, 2023. URL https://arxiv.org/abs/2311.01455.

[77] Zhiruo Wang, Daniel Fried, and Graham Neubig. Trove: Inducing verifiable and efficient toolboxes for solving programmatic tasks. *arXiv preprint arXiv:2401.12869*, 2024. URL https://arxiv.org/abs/2401.12869.

[78] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, et al. Autogen: Enabling next-gen llm applications via multi-agent conversations. In *First conference on language modeling*, 2024.

[79] Wayne Wu, Honglin He, Jack He, Yiran Wang, Chenda Duan, Zhizheng Liu, Quanyi Li, and Bolei Zhou. Metaurban: An embodied ai simulation platform for urban micromobility, 2024. URL https://arxiv.org/abs/2407.08725.

[80] Zhiheng Xi, Yiwen Ding, Wenxiang Chen, Boyang Hong, Honglin Guo, Junzhe Wang, Xin Guo, Dingwen Yang, Chenyang Liao, Wei He, et al. Agentgym: Evaluating and training large language model-based agents across diverse environments. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 27914–27961, 2025.

[81] Hongchi Xia, Xuan Li, Zhaoshuo Li, Qianli Ma, Jiashu Xu, Ming-Yu Liu, Yin Cui, Tsung-Yi Lin, Wei-Chiu Ma, Shenlong Wang, Shuran Song, and Fangyin Wei. Sage: Scalable agentic 3d scene generation for embodied ai, 2026. URL https://arxiv.org/abs/2602.10116.

[82] Tianbao Xie, Siheng Zhao, Chen Henry Wu, Yitao Liu, Qian Luo, Victor Zhong, Yanchao Yang, and Tao Yu. Text2reward: Reward shaping with language models for reinforcement learning. *arXiv preprint arXiv:2309.11489*, 2023. URL https://arxiv.org/abs/2309.11489.

[83] Tianbao Xie, Danyang Zhang, Jixuan Chen, Xiaochuan Li, Siheng Zhao, Ruisheng Cao, Toh J Hua, Zhoujun Cheng, Dongchan Shin, Fangyu Lei, et al. Osworld: Benchmarking multimodal agents for open-ended tasks in real computer environments. *Advances in Neural Information Processing Systems*, 37:52040–52094, 2024.

[84] John Yang, Carlos E Jimenez, Alexander Wettig, Kilian Lieret, Shunyu Yao, Karthik Narasimhan, and Ofir Press. Swe-agent: Agent-computer interfaces enable automated software engineering. *Advances in Neural Information Processing Systems*, 37:50528–50652, 2024.

[85] John Yang, Carlos E Jimenez, Alex L Zhang, Kilian Lieret, Joyce Yang, Xindi Wu, Ori Press, Niklas Muennighoff, Gabriel Synnaeve, Karthik R Narasimhan, et al. Swe-bench multimodal: Do ai systems generalize to visual software domains? *arXiv preprint arXiv:2410.03859*, 2024. URL https://arxiv.org/abs/2410.03859.

[86] Sherry Yang, Yilun Du, Kamyar Ghasemipour, Jonathan Tompson, Leslie Kaelbling, Dale Schuurmans, and Pieter Abbeel. Learning interactive real-world simulators. *arXiv preprint arXiv:2310.06114*, 2023. URL https://arxiv.org/abs/2310.06114.

[87] Yue Yang, Fan-Yun Sun, Luca Weihs, Eli VanderBilt, Alvaro Herrasti, Winson Han, Jiajun Wu, Nick Haber, Ranjay Krishna, Lingjie Liu, Chris Callison-Burch, Mark Yatskar, Aniruddha Kembhavi, and Christopher Clark. Holodeck: Language guided generation of 3d embodied ai environments, 2024. URL https://arxiv.org/abs/2312.09067.

[88] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022. URL https://arxiv.org/abs/2210.03629.

[89] Xiaokang Ye, Jiawei Ren, Yan Zhuang, Xuhong He, Yiming Liang, Yiqing Yang, Mrinaal Dogra, Xianrui Zhong, Eric Liu, Kevin Benavente, Rajiv Mandya Nagaraju, Dhruv Sharma, Ziqiao Ma, Tianmin Shu, Zhiting Hu, and Lianhui Qin. Simworld: An open-ended simulator for agents in physical and social worlds. In *Advances in Neural Information Processing Systems*, 2025.

[90] Xunjian Yin, Xinyi Wang, Liangming Pan, Li Lin, Xiaojun Wan, and William Yang Wang. Gödel agent: A self-referential agent framework for recursive self-improvement. *arXiv preprint arXiv:2410.04444*, 2024. URL https://arxiv.org/abs/2410.04444.

[91] Hong-Xing Yu, Haoyi Duan, Charles Herrmann, William T. Freeman, and Jiajun Wu. Wonderworld: Interactive 3d scene generation from a single image, 2025. URL https://arxiv.org/abs/2406.09394.

[92] Abhay Zala, Jaemin Cho, Han Lin, Jaehong Yoon, and Mohit Bansal. Envgen: Generating and adapting environments via llms for training embodied agents, 2024. URL https://arxiv.org/abs/2403.12014.

[93] Daoguang Zan, Zhirong Huang, Wei Liu, Hanwu Chen, Linhao Zhang, Shulin Xin, Lu Chen, Qi Liu, Xiaojian Zhong, Aoyan Li, et al. Multi-swe-bench: A multilingual benchmark for issue resolving. *arXiv preprint arXiv:2504.02605*, 2025. URL https://arxiv.org/abs/2504.02605.

[94] Guangyao Zhai, Evin Pınar Örnek, Shun-Cheng Wu, Yan Di, Federico Tombari, Nassir Navab, and Benjamin Busam. Commonscenes: Generating commonsense 3d indoor scenes with scene graph diffusion, 2023. URL https://arxiv.org/abs/2305.16283.

[95] Jenny Zhang, Shengran Hu, Cong Lu, Robert Lange, and Jeff Clune. Darwin godel machine: Open-ended evolution of self-improving agents. *arXiv preprint arXiv:2505.22954*, 2025. URL https://arxiv.org/abs/2505.22954.

[96] Qihang Zhang, Chaoyang Wang, Aliaksandr Siarohin, Peiye Zhuang, Yinghao Xu, Ceyuan Yang, Dahua Lin, Bolei Zhou, Sergey Tulyakov, and Hsin-Ying Lee. Scenewiz3d: Towards text-guided 3d scene composition. *arXiv preprint arXiv:2312.08885*, 2023. URL https://arxiv.org/abs/2312.08885.

[97] Qizheng Zhang, Changran Hu, Shubhangi Upasani, Boyuan Ma, Fenglu Hong, Vamsidhar Kamanuru, Jay Rainton, Chen Wu, Mengmeng Ji, Hanchen Li, et al. Agentic context engineering: Evolving contexts for self-improving language models. *arXiv preprint arXiv:2510.04618*, 2025. URL https://arxiv.org/abs/2510.04618.

[98] Yi Zhang, Yunshuang Wang, Zeyu Zhang, and Hao Tang. Code2worlds: Empowering coding llms for 4d world generation. *arXiv preprint arXiv:2602.11757*, 2026. URL https://arxiv.org/abs/2602.11757.

[99] Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. Expel: Llm agents are experiential learners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19632–19642, 2024.

[100] Fangwei Zhong, Kui Wu, Churan Wang, Hao Chen, Hai Ci, Zhoujun Li, and Yizhou Wang. Unrealzoo: Enriching photo-realistic virtual worlds for embodied ai, 2025. URL https://arxiv.org/abs/2412.20977.

[101] Qinhong Zhou, Hongxin Zhang, Xiangye Lin, Zheyuan Zhang, Yutian Chen, Wenjun Liu, Zunzhe Zhang, Sunli Chen, Lixing Fang, Qiushi Lyu, Xinyu Sun, Jincheng Yang, Zeyuan Wang, Bao Chi Dang, Zhehuan Chen, Daksha Ladia, Quang Vinh Dang, Jiageng Liu, and Chuang Gan. Virtual community: An open world for humans, robots, and society, 2026. URL https://arxiv.org/abs/2508.14893.

[102] Yan Zhuang, Jiawei Ren, Xiaokang Ye, Jianzhi Shen, Ruixuan Zhang, Tianai Yue, Muhammad Faayez, Xuhong He, Ziqiao Ma, Lianhui Qin, Zhiting Hu, and Tianmin Shu. Simworld-robotics: Synthesizing photorealistic and dynamic urban environments for multimodal robot navigation and collaboration, 2026. URL https://arxiv.org/abs/2512.10046.

[103] Terry Yue Zhuo, Minh Chien Vu, Jenny Chim, Han Hu, Wenhao Yu, Ratnadira Widyasari, Imam Nur Bani Yusuf, Haolan Zhan, Junda He, Indraneil Paul, et al. Bigcodebench: Benchmarking code generation with diverse function calls and complex instructions. *arXiv preprint arXiv:2406.15877*, 2024. URL https://arxiv.org/abs/2406.15877.

[104] Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, Quan Vuong, Vincent Vanhoucke, Huong Tran, Radu Soricut, Anikait Singh, Jaspiar Singh, Pierre Sermanet, Pannag R. Sanketi, Grecia Salazar, Michael S. Ryoo, Krista Reymann, Kanishka Rao, Karl Pertsch, Igor Mordatch, Henryk Michalewski, Yao Lu, Sergey Levine, Lisa Lee, Tsang-Wei Edward Lee, Isabel Leal, Yuheng Kuang, Dmitry Kalashnikov, Ryan Julian, Nikhil J. Joshi, Alex Irpan, Brian Ichter, Jasmine Hsu, Alexander Herzog, Karol Hausman, Keerthana Gopalakrishnan, Chuyuan Fu, Pete Florence, Chelsea Finn, Kumar Avinava Dubey, Danny Driess, Tianli Ding, Krzysztof Marcin Choromanski, Xi Chen, Yevgen Chebotar, Justice Carbajal, Noah Brown, Anthony Brohan, Montserrat Gonzalez Arenas, and Kehang Han. RT-2: Vision-language-action models transfer web knowledge to robotic control. In *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pages 2165–2183. PMLR, 2023.

A SIMWORLD STUDIO: Platform Details

A.1 MCP Tool Reference

Table 4 lists all 14 MCP tools exposed by the SIMWORLD STUDIO server. Tools are organized into four functional groups: actor management, environment and asset management, scene evaluation, and a Python escape hatch for operations not covered by the predefined API.

Table 4: **SIMWORLD STUDIO MCP server tool reference.** $*$ = required parameter.

Group	Tool	Purpose	Key Parameters
Actor management	`spawn_blueprint_actor`	Spawn a CityDatabase Blueprint, such as a building, tree, vehicle, or prop; accepts full path or shorthand.	`actor_name`*, `blueprint_id`*, `location`*, `rotation`, `scale`
	`spawn_actor`	Spawn a static-mesh actor, including engine primitives or any `SM_` asset.	`name`*, `static_mesh`*, `location`*, `rotation`, `scale`
	`delete_actor`	Delete a named actor.	`name`*
	`delete_all_spawned`	Delete all session-spawned actors.	—
	`get_actors_in_level`	List every actor currently in the UE level.	—
	`find_actors_by_name`	Search actors by name pattern.	`pattern`*
	`set_actor_transform`	Move, rotate, or scale an actor.	`name`*, `location`, `rotation`, `scale`
	`take_screenshot`	Capture the UE viewport as PNG.	`filename`
Environment & assets	`setup_environment`	Initialize lighting, sky, fog, ground plane, and view distance. Must be called before spawning assets.	`ground_size`, `time_of_day`
	`list_assets`	List available SimWorld assets by category, including buildings, trees, vehicles, street furniture, and roads.	`category`
Scene evaluation	`verify_scene`	VLM verifier: captures the scene and evaluates placement against the original request; returns `PASS`, `NEEDS_IMPROVEMENT`, or `FAIL` with actionable suggestions.	`original_request`*, `focus_areas`
	`check_collisions`	Geometric overlap check: computes world-space AABBs and returns intersecting pairs with area in cm^2.	`names`, `scope`, `min_area_cm2`
	`check_vertical_support`	Detects floating objects by reporting actors whose AABB bottom is above ground and not supported by another actor.	`names`, `scope`, `ground_z`, `tolerance_cm`
Python escape hatch	`execute_python_script`	Execute arbitrary Unreal Engine Python for operations not covered by dedicated tools; successful scripts can be saved as reusable skills.	`script`*

A.2 SIMCODER Generation Pipeline

Figure 2 illustrates the overall SIMCODER architecture. Here we describe each stage of the generation pipeline in detail.

Stage 1: Context acquisition. SIMCODER begins every generation episode by querying its context through tool calls rather than a manually engineered prompt. It calls `list_assets` to enumerate available mesh categories and asset counts, queries the `SkillRegistry` to retrieve applicable skills (e.g., *building-placement*, *city-layout*), and calls `setup_environment` to initialize the scene. This tool-driven context acquisition ensures that SIMCODER always operates with an accurate, up-to-date view of its action space.

Stage 2: Layout planning. Given the natural-language prompt and optional reference image, SIMCODER drafts a high-level layout plan specifying spatial zones (e.g., street grid, park, commercial row), asset categories per zone, and approximate density. For image-guided generation (S2), the plan is anchored to spatial structure inferred from the reference. Layout planning is performed in-context; the plan is not externalized but is used to sequence subsequent tool calls.

Stage 3: Iterative scene construction. SIMCODER builds the scene incrementally, spawning and arranging actors via actor-management tools. After each spawn batch, it calls `check_collisions` and `check_vertical_support`; violations are returned inline as tool-call responses and resolved

before proceeding. This per-call verification loop (§2.1) prevents error compounding across a long construction trajectory.

Stage 4: VLM-based semantic verification. After completing a logical block of construction, SIMCODER calls `verify_scene` with the original prompt. The VLM (Claude in our experiments) receives six multi-angle screenshots and the current actor list, scores semantic alignment, and returns structured feedback (PASS / NEEDS_IMPROVEMENT / FAIL) with specific issue descriptions. If the verdict is not PASS, SIMCODER performs targeted corrections and re-verifies. The generation episode terminates on PASS or after a maximum of three verification rounds.

Stage 5: Skill authoring (self-evolution). When verifier feedback reveals a failure pattern that recurs across episodes (e.g., commercial buildings colliding due to under-estimated footprints), SIMCODER authors a corrective skill or tool wrapper via `execute_python_script` and writes it to the skill directory. On the next skill-registry refresh, the new skill is indexed and becomes available to all future generations, permanently extending the agent's capability without manual intervention.

A.3 Skill Library Structure

Each skill is a Markdown document with a YAML front-matter header and an optional companion Python utility:

```
---
name: building-placement
version: 1.2
tags: [placement, spacing, buildings]
dependencies: [city-layout]
python_util: building_placement_utils.py
---
## Summary
Footprint-aware spacing rules for urban building placement ...
## Usage
Call ‘compute_building_grid(n_buildings, style)‘ to get
a list of (location, rotation) placements respecting
minimum inter-building clearance per size category.
```

The five built-in skills cover: **building-placement** (size categories, spacing tables, rotation patterns across 127 building classes), **city-layout** (grid blocks, street-facing rows, mixed-use neighborhoods), **street-furniture** (placement heuristics for trees, benches, lamps, and signs), **weather-and-mood** (lighting and time-of-day presets), and **screenshot-tour** (multi-angle capture patterns for the VLM verifier). Self-authored skills follow the same schema, ensuring seamless integration into the registry.

A.4 Detailed Running Case of SIMCODER

This section provides a detailed running case of SIMCODER editing a container-yard maze scene through iterative tool use, feedback injection, and skill accumulation. The task is to build a navigable container maze near `PlayerStart`. Across four shown rounds, SIMCODER incrementally constructs the maze entrance, extends it into a T-junction, adds a diagonal obstacle, and introduces dead-end formations. The case illustrates three core mechanisms of SIMWORLD STUDIO: external verifier feedback, collision-aware correction, and reusable skill formation.

Running Case Overview: Container-Yard Maze Editing

Run ID. `20260505_013558`
Task. Build a navigable container-yard maze near `PlayerStart` using iterative scene editing.
Model. Claude Sonnet 4.5 via Claude Code SDK.
Rounds shown. 4.
Skills learned in full run. 6.
Final recorded VLM score. 6.5/10 at Round 3.

Each round follows a fixed loop. The agent receives scene context and feedback from the previous round, plans the next spatial edit, invokes scene-editing tools, and emits newly learned skills when a reusable pattern is discovered. After the agent round ends, the pipeline captures screenshots, runs AABB collision checking, performs VLM evaluation in a separate SDK session, and injects the resulting feedback into the next prompt.

Pipeline Loop

Agent side: Claude SDK	**Pipeline side: subprocess**
1. Call `get_scene_summary`.	Inject VLM and collision feedback from the previous round into the prompt.
2. Apply prior VLM/collision fixes.	After the round ends, call `take_screenshot`.
3. Plan the next maze element.	Run AABB collision checking over spawned actors.
4. Call `spawn_actor` two to three times.	Run VLM evaluation in a separate SDK session.
5. Output JSON with `new_skills`.	Inject screenshot, collision, and VLM results into the next round.

Skill Library: Pre-seeded and Learned Skills

Source	**Skill**	**Description**
Pre-seeded	`get_scene_summary`	Get `PlayerStart` location and compact summaries of `Demo_` actors.
Pre-seeded	`search_ue_assets`	Search the UE asset registry by keyword for mesh paths.
Pre-seeded	`spawn_container_row`	Spawn a row of containers with configurable spacing and yaw.
Pre-seeded	`collision_check_demo_actors`	Run AABB collision checking over actors with the `Demo_` prefix.
Round 1 learned	`spawn_maze_l_junction`	L-shaped junction with two parallel walls and one perpendicular blocker.
Round 2 learned	`spawn_perpendicular_blocker`	Container at 90° yaw to close a corridor end.
Round 3 learned	`spawn_tjunction_maze_segment`	T-junction with parallel corridors and a diagonal obstacle.
Later learned	`spawn_maze_t_junction`	T-junction using perpendicular and lengthwise containers.
Later learned	`spawn_maze_dead_end`	U-shaped dead-end with one back wall and two side walls.
Later learned	`spawn_dead_end_u_formation`	Three-container U dead-end with configurable offsets.

Round 1: Maze Entrance Corridor



Agent reasoning

"I will place two parallel lengthwise containers as corridor walls and one perpendicular container as the first turn blocker. Using rotation=[0,0,0] for walls and rotation=[0,0,90] for the turn—this establishes the maze entrance."

Objects spawned.

Actor name	Mesh	Yaw	Purpose
`Demo_MazeWall_Left_XXXX`	`SM_Container12m`	0°	Left corridor wall.
`Demo_MazeWall_Right_XXXX`	`SM_Container12m`	0°	Right corridor wall.
`Demo_MazeTurn_Block_XXXX`	`SM_Container12m`	90°	First turn blocker.

Skill saved

`spawn_maze_l_junction`: spawn an L-shaped junction with two parallel walls and one perpendicular blocker.

VLM verifier feedback

Score: 4.5/10. The verifier finds a good foundation with rotation variety and industrial realism, but the maze is still too sparse. Only three containers are visible, so the structure reads as a simple corridor rather than a recognizable maze. The next round should add branching paths and small navigation obstacles.

Collision check. Not triggered in Round 1 because there are no prior maze actors to conflict with.

Saved Python Utility: `spawn_maze_l_junction`

```
# spawn_maze_l_junction
import unreal, json, time

MESH = "/Game/ContainerYard/Meshes/SM_Container12m.SM_Container12m"
BASE_X, BASE_Y, Z = -3500, -8000, 0
SPACING = 900  # corridor width

subsys = unreal.get_editor_subsystem(unreal.EditorActorSubsystem)
mesh_obj = unreal.load_asset(MESH)
```

```
configs = [
    ("Demo_WallL", BASE_X, BASE_Y, 0),
    ("Demo_WallR", BASE_X, BASE_Y + SPACING, 0),
    ("Demo_Turn", BASE_X + 1300, BASE_Y + SPACING // 2, 90),
]

spawned = []
for name, x, y, yaw in configs:
    sfx = hex(int(time.time() * 1000) % 0xFFFF)[2:].upper()
    actor = subsys.spawn_actor_from_class(
        unreal.StaticMeshActor,
        unreal.Vector(x, y, Z)
    )
    if actor:
        actor.set_actor_label(f"{name}_{sfx}")
        comp = actor.get_component_by_class(unreal.StaticMeshComponent)
        if comp:
            comp.set_static_mesh(mesh_obj)
        actor.set_actor_rotation(unreal.Rotator(yaw=yaw), False)
        spawned.append(f"{name}_{sfx}")

print(json.dumps({"spawned": spawned}))
```

Round 2: T-Junction Extension



injected context

VLM Score: 4.5/10. Add branching paths, such as a T-junction or dead-end fork, and include small navigation obstacles.

Agent reasoning

"VLM says density is too low. I will extend the corridor into a T-junction by adding two more lengthwise walls parallel to the existing corridor, plus a perpendicular blocker at the junction to force the turn decision."

Objects spawned.

Actor name	Mesh	Yaw	Purpose
`Demo_WallWest_XXXX`	`SM_Container12m`	0°	West branch wall.
`Demo_WallEast_XXXX`	`SM_Container12m`	0°	East branch wall.
`Demo_TjuncBlock_XXXX`	`SM_Container12m`	90°	T-junction blocker.
`Demo_TjuncBlock_XXXX`	`SM_Container12m`	90°	Retry spawn to confirm junction closure.

Skill saved

`spawn_perpendicular_blocker`: spawn a container at 90° yaw to close a corridor end or force a turn.

VLM verifier feedback

Score: 5.5/10. The verifier confirms visible rotation variety and a clearer L-turn geometry. However, the scene is still closer to a simple hallway than a branching maze. The next round should add a U-shaped dead end, close the north gap with a perpendicular blocker, and add small props inside corridors.

Collision check. One collision pair is detected between `Demo_MazeWall_01` and `Demo_EntranceWall`. The agent fixes this by deleting the overlapping actor before continuing.

Round 3: Diagonal Obstacle and North Extension



injected context

VLM Score: 5.5/10. Add a U-shaped dead end, close the north gap with a perpendicular blocker, and spawn two to three small props inside corridors.

Agent reasoning

"I will extend north with two more parallel walls, add a 45-degree diagonal container as a navigation obstacle inside the corridor to break straight-line visibility, saving as a T-junction skill for future reuse."

Objects spawned.

Actor name	Mesh	Yaw	Purpose
`Demo_NorthWallWest_XXXX`	`SM_Container12m`	0°	North corridor wall.
`Demo_NorthWallEast_XXXX`	`SM_Container12m`	0°	North corridor wall.
`Demo_DiagObstacle_XXXX`	`SM_Container12m`	45°	Diagonal obstacle inside corridor.

Skill saved

`spawn_tjunction_maze_segment`: spawn a T-junction maze segment with parallel corridors and a diagonal obstacle.

VLM verifier feedback

Score: 6.5/10. The verifier confirms rotation diversity through diagonal container placement and a wider branching footprint. Navigable aisles remain visible and the industrial context is preserved. The scene is still below full maze complexity, so future rounds should add more containers, dead-end paths, and small interior props.

Collision check. No collisions are detected.

Round 4: Dead-End Formations



Injected context

VLM Score: 6.5/10. Add U-shaped dead-end formations and close open gaps.

Agent reasoning

"I will build two U-shaped dead-ends to force the agent into backtracking decisions. This round uses six spawns, which is above the usual limit, but it is useful for increasing maze complexity."

Objects spawned.
Six containers are spawned to form two U-shaped dead-end structures. The formation uses lengthwise walls at 0° yaw and perpendicular blockers at 90° yaw.

VLM verifier feedback

Round 4 VLM feedback is not recorded because of a pipeline screenshot path error.

Collision check. No collisions are detected.

Score Progression

Round	VLM Score	Collisions	Skills Learned	Key Action
1	4.5/10	0	1	Entrance corridor with L-shaped turn.
2	5.5/10	$1 \rightarrow 0$	1	T-junction extension and collision repair.
3	6.5/10	0	1	North extension with 45° diagonal obstacle.
4	—	0	0	Dead-end formations for backtracking behavior.

Evidence of Pipeline Components

Skill reuse. Later rounds explicitly retrieve and reuse earlier skills such as `spawn_perpendicular_blocker` and `spawn_maze_l_junction`, showing that learned scene-editing patterns are persisted across rounds.

VLM-to-agent feedback loop. The VLM feedback is injected into the next prompt as hard context. For example, after Round 1, the agent receives a score of 4.5/10 and the instruction to add branching paths. The next response directly acts on this feedback by extending the corridor into a T-junction.

Collision-aware auto-fix. When the pipeline detects a collision between `Demo_MazeWall_01` and `Demo_EntranceWall`, the next prompt requires the agent to delete or move the overlapping actor before adding new objects. This turns geometric validation into an explicit correction signal.

Rotation diversity. The run gradually increases geometric complexity by using 0°, 45°, 90°, and later 135° yaw placements. This produces corridors, perpendicular blockers, diagonal obstacles, and dead-end formations.

Level Collision Checking

The collision checker runs after every round over actors with the `Demo_` prefix. It performs a lightweight AABB overlap test and injects detected collisions into the next prompt as mandatory correction context.

Level AABB Collision Check

```
# Pipeline-level AABB collision check.
for i in range(len(demo_actors)):
    for j in range(i + 1, len(demo_actors)):
        a, b = demo_actors[i], demo_actors[j]
        if (abs(a.x - b.x) < a.ex + b.ex and
            abs(a.y - b.y) < a.ey + b.ey):
            collision_pairs.append({
                "a": a.label,
                "b": b.label
            })
```

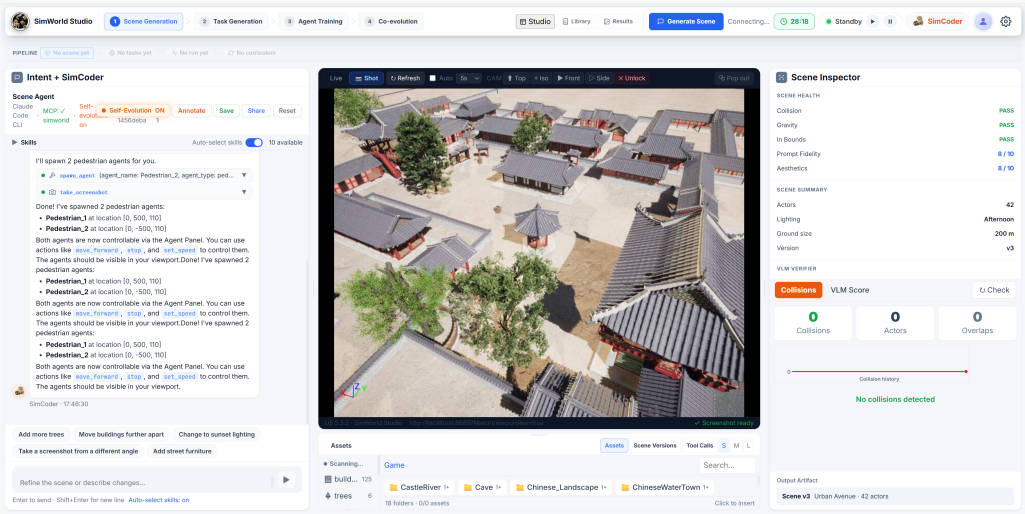
The resulting message is injected into the next round as mandatory correction context:

```
COLLISIONS: N pair(s) detected -> DELETE or MOVE before spawning.
```

A.5 SIMWORLD STUDIO's Graphic User Interface

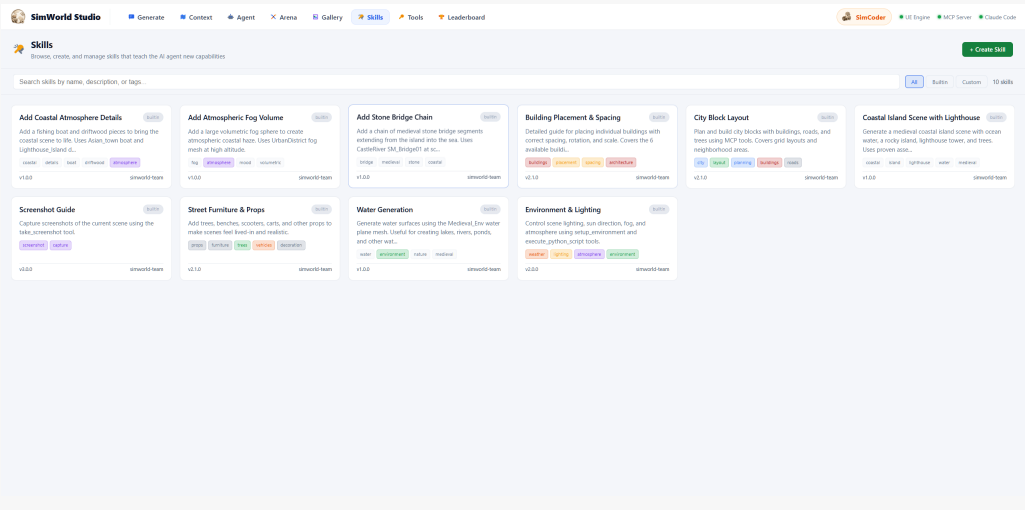
Figure 8 presents representative interface views of SIMWORLD STUDIO. The first panel shows the main studio interface, integrating the generation pipeline, user–agent interaction, UE rendering, an asset browser, and a scene inspector. The remaining dark-theme panels show specialized views for skill management, tool abstraction, and interactive user control inside the generated Unreal Engine environment.

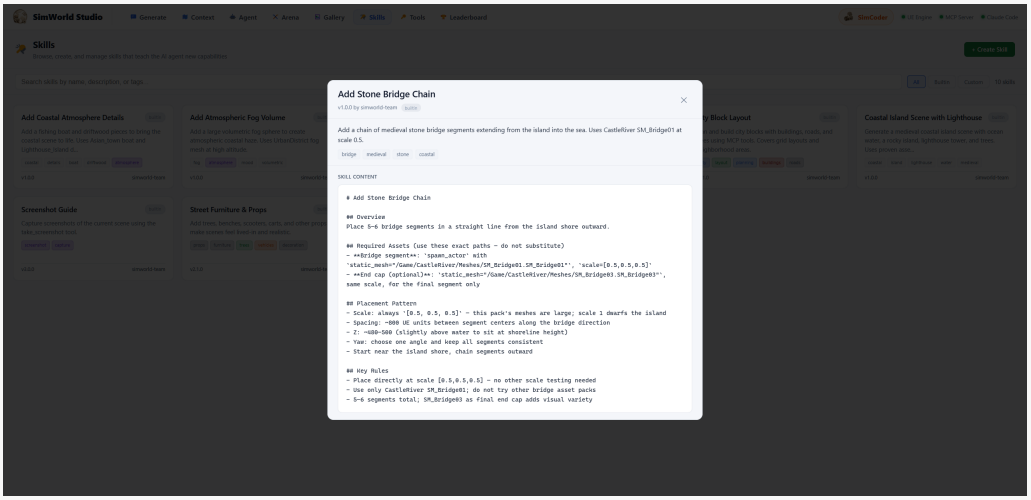
Appendix Case A: Main SIMWORLD STUDIO Interface



(a) Main studio interface. The light-theme main page integrates user–agent chat, the live UE-rendered environment, asset and backend services, Gym APIs, and embodied-agent monitoring in a single workspace. The left panel supports prompt-based interaction with SIMCODER; the center viewport displays the generated Unreal Engine scene and task targets; the bottom panel exposes assets, MCP connectivity, and feedback channels; and the right panel monitors RGB-D observations, task generation, Python APIs, and embodied-agent performance.

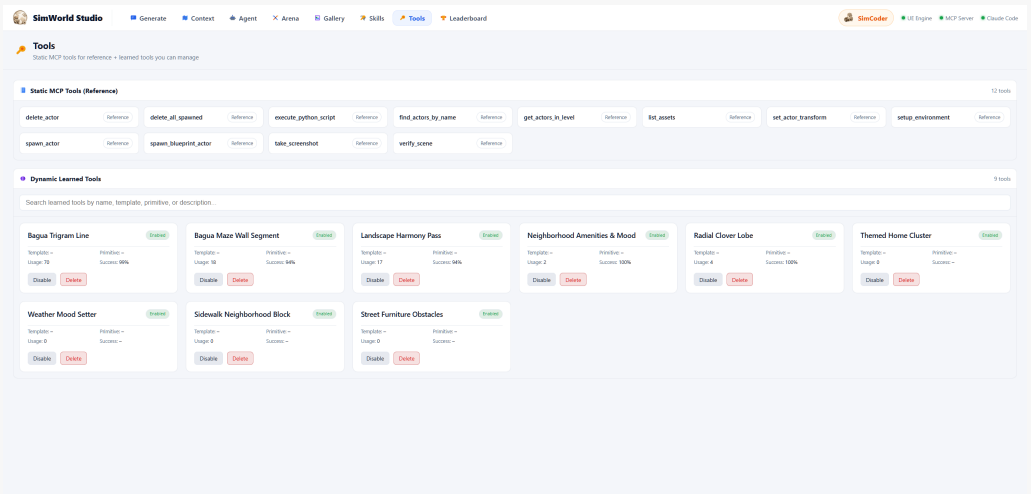
Appendix Case B: Skills Panel





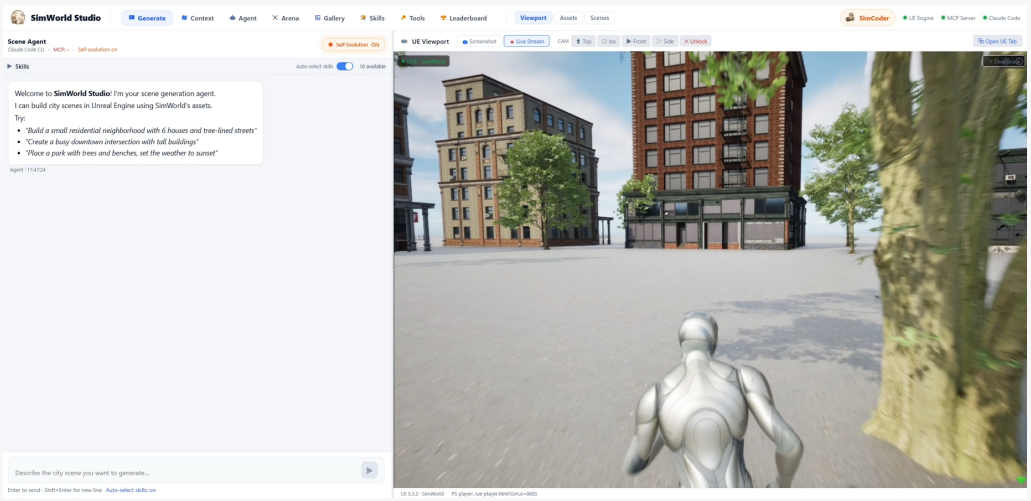
(b) Skills panel. Reusable high-level skills are organized as compositional capabilities that can be browsed, created, and reused by the coding agent across scene-generation episodes. The detailed view further exposes the internal structure of skill `add_stone_bridge_chain`, including its description, callable interface, implementation details, and execution history, enabling the agent to inspect and reuse learned abstractions across tasks.

Appendix Case C: Tools Panel



(c) Tools panel. The interface exposes both static MCP tools and dynamically learned tools, enabling the agent to combine primitive operations with reusable abstractions.

Appendix Case D: Interactive User-Control Mode



(d) Interactive user-control mode. Beyond text-based prompting, users can directly control an embodied character with keyboard input to explore, inspect, and test the generated Unreal Engine environment from an egocentric perspective.

Figure 8: **Representative interface views of SIMWORLD STUDIO.** The light-theme main interface provides an integrated workspace for user–agent interaction, UE scene rendering, asset/backend management, Gym environment APIs, and embodied-agent monitoring. The dark-theme panels further show specialized views for skill management, tool abstraction, and direct embodied interaction, allowing users to move beyond text-only prompting and interact with generated environments through controllable agents.

A.6 SIMWORLD STUDIO Running Configuration

Table 5: SIMWORLD STUDIO Minimal System Requirements.

Component	Requirement
OS	Linux; Ubuntu 20.04+ recommended
GPU	NVIDIA GPU with 8GB+ VRAM; tested on L40S, T4, and A100
NVIDIA Driver	Version 525+ with Vulkan support
Node.js	Version 18+
Python	Version 3.9+
Disk Space	40 GB free; 15 GB download + 21 GB extracted

A.7 Limitations

We presented SIMWORLD STUDIO, a platform that connects coding-agent-driven 3D scene construction with embodied agent training in Unreal Engine. By exposing structured tools for scene editing, validation, and environment export, SIMWORLD STUDIO enables agents to generate diverse interactive environments and convert them into standardized embodied learning tasks. Our results show that this workflow can improve the scalability of environment construction and provide a practical path toward adaptive embodied-agent training.

A current limitation is that the effectiveness of SIMWORLD STUDIO is still bounded by the capability of the underlying coding agent. In particular, generating complex 3D environments requires strong spatial reasoning: the agent must understand object placement, geometric constraints, physical plausibility, navigability, and long-range layout consistency. While the current system can already produce useful scenes, failures may still occur when the task requires fine-grained spatial planning or precise multi-object arrangement. Improving spatial reasoning for coding agents is therefore an important direction for future work. All evaluated transfer is between simulated environments; appearance, dynamics, sensing, and actuation gaps to physical robots remain unmeasured. Additional

limitations of evaluation scale, judge validation, and the manipulation study are discussed in §5 and Appendix F.

A.8 Broader Impact

SIMWORLD STUDIO has the potential to improve productivity in 3D environment creation, especially for complex engines such as Unreal Engine where existing coding-agent support remains limited. By allowing coding agents to directly edit, validate, and reuse scene-building skills, the system can reduce repetitive engineering effort and make interactive environment construction more accessible to researchers and developers. It also provides a scalable route for embodied AI research, where the lack of diverse, controllable, and interactive environments is often a major bottleneck.

At the same time, more capable automated scene-generation tools may affect existing workflows in game development, simulation design, and digital-content production. Some routine environment-authoring tasks could become increasingly automated, which may shift the role of human creators from manual construction toward supervision, design specification, and quality control. We believe such systems should be developed as assistive tools that augment human creativity and engineering productivity, while preserving human oversight over artistic direction, safety, and deployment decisions.

Potential misuse includes generating restricted, unsafe, or policy-violating simulated environments, using automated scene construction for surveillance-like or tactical planning scenarios, or recombining licensed assets outside their permitted terms. Our release is intended for research use. We do not release trained harmful policies, scraped personal data, or human-subject datasets. The Python execution interface is local/self-hosted, and users are expected to follow the licenses and deployment policies of the engine, model providers, and asset libraries.

A.9 Assets, Licenses, and Model Access

Table 6: Existing software, model APIs, and assets used in SIMWORLD STUDIO.

Asset / Software	Source / Owner	Version / Access	License / Terms
Unreal Engine	Epic Games	UE5.3.2	Unreal Engine EULA
SimWorld asset library	SimWorld[89]	Project asset library	Follows the original SimWorld asset license / terms
Qwen3.5-2B/9B/27B	Qwen Team	Hugging Face model cards	Apache-2.0 / model card terms
Claude Opus / Sonnet APIs	Anthropic	API model snapshots used in experiments	Anthropic API terms and model documentation
Claude Code	Anthropic	Agentic coding framework	Anthropic product / API terms
Gymnasium and Python packages	Upstream open-source projects	Listed in repository environment file	Respective open-source licenses

We do not scrape personal data or release human-subject data. Generated scenes are created from prompts and licensed engine/assets. Users of the released code are expected to comply with the licenses and terms of the underlying engine, model providers, and asset libraries.

B Additional Related Work

3D scene and world generation. *Generative 3D methods* synthesize explicit 3D scene representations such as object layouts, meshes, or 3D Gaussian splats. Data-driven indoor methods model object arrangements through autoregressive, diffusion-based, or scene-graph-conditioned generation [52, 67, 94, 9], while 2D-prior-based methods construct scenes through depth estimation, inpainting, and multi-view fusion [29, 13, 91, 96]. *Video-based world models* instead treat world generation as conditioned video synthesis, producing controllable visual experiences without exposing editable 3D assets, persistent scene graphs, or physics-simulator state [10, 86, 49, 31]. *Procedural and agentic pipelines* generate simulation-ready environments through procedural generators or LLM/VLM-driven agents that combine layout planning, asset retrieval or synthesis, placement, and iterative validation [14, 59, 87, 33, 81, 53, 39, 98, 45]. SimWorld Studio targets high-fidelity open-ended world generation in Unreal Engine 5, provides a standardized gym-like interface for embodied-agent training, and incorporates a persistent verifier-driven skill library for continual improvement of the coding agent.

Tool-augmented coding agents. Large language models have evolved from passive code generators into agents that reason, invoke tools, execute code, and interact with external environments. Recent work has developed this paradigm through interleaved reasoning and acting, large-scale API invocation and orchestration, executable code actions, and reflective feedback [88, 62, 57, 74, 56, 11, 65]. Evaluation has shifted from function-level code synthesis [12, 7] to compositional library use [103], repository-level patch generation with executable test harnesses [37, 93, 85], and holistic agent evaluation across operating systems and interactive environments [43, 83]. Another line develops reusable agent infrastructure, including executable skill libraries, generated tools, role-specialized agents, agent–computer interfaces, sandboxed harnesses, and multi-agent coordination frameworks for reproducible verifier-guided iteration [70, 11, 77, 30, 78, 84, 75, 80]. SimWorld Studio extends tool-augmented coding agents beyond software repair and digital automation to verified game-engine construction of simulation-ready embodied environments.

Self-evolving agents. Recent work on self-evolving agents studies closed-loop systems that autonomously improve from environmental feedback, self-generated curricula, and agent–environment co-evolution [25, 21]. In language and software domains, key mechanisms include verbal self-reflection, experience accumulation, self-editing code, and automated search over agent designs [65, 48, 99, 90, 32, 60, 97, 95]. In web and general interactive settings, agents improve through failure-driven online curricula, multi-turn reinforcement learning, or iterative co-evolution with world models [55, 28, 22, 80, 16]. In embodied and robotic settings, self-evolution appears as adaptive training loops that expand an agent's skills and training distribution through automatic curricula, executable skill libraries, and generated tasks, rewards, or environments [70, 19, 76, 47, 82, 42, 46]. SimWorld Studio is distinct in that it self-evolves the 3D world generator itself, guided by executable scene verifiers and downstream embodied-agent usability feedback, rather than primarily improving the agent, policy, or task curriculum.

Agent–environment co-evolution. In unsupervised environment design, environments are generated, selected, or edited near the agent's capability frontier to induce automatic curricula [15, 35, 51], with open-ended and multi-agent variants jointly adapting environments, solvers, or co-players [72, 73, 61]. LLM-based methods move co-evolution into code space, using language models to generate and iteratively revise environment configurations, terrains, rewards, or task programs from agent feedback or learnability signals [92, 47, 19, 42]; GenEnv [26] further co-trains the agent and a generative environment policy through RL difficulty-aligned curriculum. A parallel world-model line uses learned simulators for VLA and RL policy improvement and iteratively co-refines the world model and policy from imagined rollouts, real interactions, or policy failures [36, 64, 44, 27, 68]. R-Zero demonstrates an analogous Challenger–Solver loop in which task generation and reasoning ability co-evolve without human-curated data [34]. SimWorld Studio differs by generating simulation-ready 3D UE5 worlds via code and adapting future worlds from downstream embodied-agent feedback.

Table 7: Comparison with representative embodied simulation platforms, generative scene-construction systems, and environment-agent co-evolution methods. **Diverse Gen.**: supports diverse environment generation at scale. **Phys./Vis. Realism**: fidelity of physics simulation and visual rendering. **Gym Interface**: provides a standard Gymnasium-compatible API. **Self-Evolution**: autonomous improvement of the platform's own generation or operation behavior from verifier feedback. **Co-Evolution**: environment generation adapts based on downstream embodied agent performance.

Platform	Engine	Diverse Gen.	Phys./Vis. Realism	Gym Interface	Self-Evolution	Co-Evolution
CARLA [17]	UE4	✗	+++	✗	✗	✗
ThreeDWorld [23]	Unity	✗	++	✗	✗	✗
AI2-THOR [38]	Unity	✗	++	✗	✗	✗
MineDojo [20]	Minecraft	✗	+	✓	✗	✗
ProcTHOR [14]	Unity	✓	++	✗	✗	✗
Habitat 3.0 [54]	Habitat-Sim	✗	++	✓	✗	✗
MetaUrban [79]	PyBullet	✓	++	✓	✗	✗
GRUtopia [71]	Isaac Sim	✓	++	✓	✗	✗
EmbodiedCity [24]	UE4	✗	+++	✗	✗	✗
UnrealZoo [100]	UE4/5	✗	+++	✓	✗	✗
Virtual Community [101]	Genesis	✓	++	✗	✗	✗
VirtualEnv [66]	UE5	✗	+++	✗	✗	✗
Holodeck [87]	AI2-THOR/Unity	✓	++	✗	✗	✗
SAGE [81]	Isaac Sim	✓	+++	✗	✓	✗
GenEnv [26]	AlfWorld/Text-only	✗	+	✗	✗	✓
SIMWORLD STUDIO	**UE5**	✓	+++	✓	✓	✓

C Case Study 1

C.1 Full Results by Difficulty Level

Table 8 reports per-metric performance broken down by difficulty level (easy/medium/hard) for all four LLM backbones and three generation settings.

Table 8: **Scene generation quality across LLM backbones, settings, and difficulty levels.** All metrics $\in [0, 1]$ (↑). Bold = best per column within each setting. E/M/H = easy/medium/hard; Avg rows average each metric across difficulties; no cross-metric aggregate is reported. Metric colors: quantity, physical validity, semantic, aesthetic.

		S1: Text-to-Scene						S2: Image+Text-to-Scene						S3: Scene Editing					
		Rule-Based				*VLM*		*Rule-Based*				*VLM*		*Rule-Based*				*VLM*	
LLM	**Diff.**	*Count*	*Diversity*	*No Collision*	*Gravity*	*Fidelity*	*Aesthetics*	*Count*	*No Collision*	*Gravity*	*In-Bounds*	*Fidelity*	*Style*	*Preserve*	*Edit Count*	*No Collision*	*Coherence*	*Edit Compl.*	*Layout*
Qwen 9B	E	.00	.70	.60	.80	.30	.20	.80	.50	.70	1.0	.20	.20	1.0	.00	.60	.30	.30	.30
	M	.00	.50	.30	.50	.20	.10	.60	.30	.50	1.0	.10	.10	1.0	.00	.30	.23	.20	.20
	H	.00	.30	.20	.43	.10	.10	.40	.20	.40	1.0	.10	.10	1.0	.00	.20	.17	.20	.10
	Avg	.00	.50	.37	.58	.20	.13	.60	.33	.53	1.0	.13	.13	1.0	.00	.37	.23	.23	.20
Qwen 27B	E	.50	.60	1.0	1.0	.60	.50	1.0	.95	1.0	1.0	.50	.37	1.0	.00	.99	.50	.20	**.60**
	M	.00	.46	.99	1.0	.40	.30	.70	.80	.90	1.0	.37	.30	1.0	.00	.99	.47	.10	.50
	H	.00	.32	.98	1.0	.30	.30	.50	.50	.60	1.0	.27	.23	1.0	.00	.99	.43	.10	.40
	Avg	.17	.46	.99	1.0	.43	.37	.73	.75	.83	1.0	.38	.30	1.0	.00	.99	.47	.13	.50
Sonnet 4	E	1.0	**1.0**	1.0	1.0	.70	.50	1.0	**.99**	**1.0**	1.0	.63	.47	1.0	1.0	.99	.50	**.90**	.50
	M	.00	**.80**	**1.0**	1.0	.50	.30	.80	**.85**	.90	1.0	.50	.37	1.0	1.0	.98	.40	**.60**	.40
	H	.00	.60	**1.0**	1.0	**.40**	**.30**	.60	.60	**.70**	1.0	.37	.27	1.0	1.0	.97	.30	**.50**	.30
	Avg	.33	**.80**	**1.0**	1.0	.53	.37	.80	.81	.87	1.0	.50	.37	1.0	1.0	.98	.40	**.67**	.40
Opus 4	E	**1.0**	.80	1.0	1.0	**.80**	**.60**	**1.0**	.99	1.0	1.0	**.80**	**.60**	1.0	1.0	.99	**.60**	.70	.60
	M	**1.0**	.72	.99	1.0	**.70**	**.50**	**.80**	.90	.95	1.0	**.60**	**.47**	1.0	1.0	.98	**.50**	.50	**.53**
	H	**.25**	**.64**	.98	1.0	.40	.30	**.70**	**.70**	.80	1.0	**.47**	**.33**	1.0	1.0	.97	**.40**	.30	**.47**
	Avg	**.75**	.72	.99	1.0	**.63**	**.47**	**.83**	**.86**	**.92**	1.0	**.62**	**.47**	1.0	1.0	.98	**.50**	.50	**.53**

C.2 Detailed Metric Specifications

Table 9 provides the complete metric inventory across all three evaluation settings.

Table 9: Complete metric inventory. "R" = Rule-based, "V" = VLM-as-Judge. **Bold** indicates metrics novel to this work.

Metric	Type	S1	S2	S3	Adapted From
CNT (Object Count)	R	✓	✓	—	SceneEval, Holodeck
DIV (Diversity)	R	✓	—	—	**Novel**
COL (Collision Rate)	R	✓	✓	✓	SAGE, VULCAN, SceneEval
GRAV (Gravity Validity)	R	✓	✓	—	SceneEval, VULCAN
OOB (In-Bounds Rate)	R	✓	—	—	SceneEval
PRES (Preservation)	R	—	—	✓	Adapted
ECNT (Edit Count)	R	—	—	✓	**Novel**
PF (Prompt Fidelity)	V	✓	✓	—	Code2Worlds
SRF (Spatial Fidelity)	V	✓	—	—	SceneEval
LAES (Aesthetics)	V	✓	—	—	WorldCraft, SAGE
ILC (Image Correspondence)	V	—	✓	—	Adapted
STY (Style Consistency)	V	—	✓	—	Adapted
EC (Edit Completeness)	V	—	—	✓	**Novel**
SC (Scene Coherence)	V	—	—	✓	**Novel**
LQ (Layout Quality)	V	—	—	✓	**Novel**

VLM scoring rubric. All VLM metrics use a 0–10 integer rubric with anchored descriptors: **10** = perfect realization; **7** = most elements correct, minor deviations; **5** = partial match, significant deviations; **3** = vaguely related, major elements missing; **0** = no relationship. Scores are normalized to $[0, 1]$ by dividing by 10 for aggregation. The VLM receives the original prompt, six multi-angle screenshots, and the scene graph as context.

Rule-based metric details. **Collision detection** uses axis-aligned bounding box (AABB) overlap tests on all non-environment actor pairs where both actors have extent > 100 units. **Gravity checking** in this evaluation is a ground-proximity proxy, distinct from the support-aware runtime verifier: it tests whether each actor's bounding box bottom face lies within ± 200 units of the ground plane ($z = 0$). **In-bounds checking** tests $|x|, |y| \leq 9500$ units (the 190m $\times$ 190m ground plane).
AABB overlap over-approximates complex or rotated meshes and can produce false positives. The ground-proximity score can also penalize valid stacked objects; neither proxy alone establishes contact-level physical realism.

C.3 Ablation Results

Table 10: Ablation study on verification loop and self-evolving skill accumulation across four models. All methods use the same minimal system prompt without domain-specific hints. **Bold** = best per model–difficulty group. Green = improvement ≥ 0.10 over baseline; red = degradation ≥ 0.10.

Model	Method	Task	CNT↑	DIV↑	COL↑	PF↑	SRF↑	LAES↑	Tools
Qwen3.5-9B	Baseline	Easy	**1.00**	**1.00**	**1.00**	0.40	0.40	0.30	8
		Mid	**1.00**	**0.90**	0.92	0.40	0.40	0.40	21
		Hard	0.00	0.40	**1.00**	0.00	0.00	0.00	10
	+ Verify	Easy	**1.00**	0.43	**1.00**	**0.60**	**0.60**	**0.40**	8
		Mid	**1.00**	0.85	0.94	0.40	0.30	0.20	21
		Hard	**0.25**	0.54	**1.00**	0.20	0.20	0.30	129
	+ Self-Evolve	Easy	**1.00**	**1.00**	**1.00**	0.50	0.40	0.30	8
		Mid	**1.00**	0.86	0.92	**0.50**	0.40	**0.50**	22
		Hard	0.00	**1.00**	**1.00**	**0.10**	**0.10**	**0.20**	27
Qwen3.5-27B	Baseline	Easy	**1.00**	**1.00**	0.80	0.50	0.30	0.30	8
		Mid	**1.00**	**0.81**	0.99	0.40	0.40	0.30	22
		Hard	0.25	**0.50**	0.88	0.20	0.20	0.30	26
	+ Verify	Easy	**1.00**	**1.00**	0.80	0.50	0.50	0.30	18
		Mid	**1.00**	0.77	0.99	**0.50**	**0.60**	**0.50**	23
		Hard	**0.25**	0.38	**0.99**	**0.40**	**0.40**	**0.50**	37
	+ Self-Evolve	Easy	**1.00**	**1.00**	0.80	0.40	0.40	0.30	8
		Mid	**1.00**	**0.81**	**1.00**	**0.50**	0.50	0.40	22
		Hard	**0.25**	**0.50**	0.95	0.30	0.30	0.40	31
Sonnet 4	Baseline	Easy	**1.00**	**1.00**	0.67	0.40	0.20	0.20	10
		Mid	**1.00**	**0.79**	**1.00**	0.60	0.50	0.50	27
		Hard	**0.50**	**0.44**	**0.99**	0.40	0.30	0.30	60
	+ Verify	Easy	0.00	**1.00**	**1.00**	0.40	0.30	0.30	21
		Mid	**1.00**	0.65	**1.00**	**0.60**	**0.60**	0.50	29
		Hard	0.25	0.36	**0.99**	**0.60**	**0.50**	**0.60**	73
	+ Self-Evolve	Easy	**1.00**	**1.00**	0.80	**0.50**	**0.50**	**0.40**	10
		Mid	**1.00**	0.78	**1.00**	0.50	0.50	0.50	26
		Hard	0.25	0.47	**0.99**	0.50	0.40	**0.50**	58
Opus 4	Baseline	Easy	**1.00**	**1.00**	0.80	0.40	0.40	0.30	11
		Mid	**1.00**	**0.85**	0.91	0.30	0.20	0.20	23
		Hard	0.25	0.37	**1.00**	0.40	0.30	0.30	89
	+ Verify	Easy	**1.00**	0.88	**1.00**	**0.70**	**0.80**	**0.50**	12
		Mid	**1.00**	0.77	0.98	**0.60**	**0.70**	**0.60**	24
		Hard	0.25	0.31	**1.00**	**0.60**	**0.50**	**0.60**	97
	+ Self-Evolve	Easy	**1.00**	**1.00**	0.80	0.00	0.40	0.40	11
		Mid	**1.00**	0.76	**1.00**	**0.60**	0.50	0.50	28
		Hard	**0.50**	**0.42**	**1.00**	0.50	0.50	0.50	62

Metrics: CNT = object count accuracy, DIV = asset diversity, COL = collision avoidance, PF = prompt fidelity (VLM), SRF = spatial relation fidelity (VLM), LAES = layout aesthetics (VLM). All $\in [0, 1]$. *Tools* = total MCP tool calls across all rounds. Self-evolving runs tasks sequentially (easy→mid→hard), accumulating skills between tasks. Verification uses up to 3 harness-driven fix rounds per task.

D Case Study 2: Experimental Details

D.1 Compute Setup for Case Study 2

For Case Study 2, we use an 8×H100 server to run the embodied-agent rollout pipeline. We host eight SIMWORLD STUDIO instances in parallel, each responsible for executing a subset of the navigation episodes. The instances are synchronized with the agent-inference workers for observation collection, action execution, reward computation, and trajectory logging. With this setup, processing the 1.2K-episode training set requires approximately 8 hours per full pass.

D.2 Experiment Design

Environment generation. Training environments are generated by SIMCODER using text-to-scene prompts spanning five urban archetypes: downtown intersections, residential neighborhoods, industrial districts, commercial avenues, and mixed-use blocks. Each environment is generated on a fixed $190\text{m} \times 190\text{m}$ ground plane and automatically exported as a Gymnasium environment via the SIMWORLD STUDIO embodied interface (§2.1). To study the effect of environmental diversity on downstream learning, we vary the number of distinct training environments from 1 to 30 while holding the total training-episode budget fixed at 200 episodes.

Episode generation. Navigation episodes are generated automatically from each environment's UE5 NavMesh via the following procedure:

1. **NavMesh computation.** UE5's NavMesh is built from the static geometry of the generated scene, producing a walkable region that inherits all obstacles and building footprints automatically.
2. **Start–goal sampling.** Start and goal positions are sampled uniformly from the walkable region. For PointNav, the goal is a 2D coordinate; for ObjectNav, the goal is a semantic category drawn from objects present in the scene.
3. **Path computation.** The geodesic shortest path between start and goal is computed via NavMesh A* search, recording the path waypoints and length L^*.
4. **Filtering.** An episode is retained if: (i) the path is fully reachable, (ii) the path length satisfies $3\text{m} \leq L^* \leq 20\text{m}$, and (iii) for ObjectNav, the target object is visible from at least one waypoint along the shortest path.

We generate 1,200 training episodes distributed across all training environments and 329 held-out test episodes on unseen SIMWORLD STUDIO-generated maps.

Agent observation and action space. At each step the agent receives a structured observation tuple: an RGB image (224×224), bearing to goal (degrees), geodesic distance to goal (meters), and elapsed step count. The discrete action space consists of four primitives: `move_forward` (0.25 m), `turn_left` ($15°$), `turn_right` ($15°$), and `stop`. An episode terminates on `stop` or after 40 steps.

Training protocol. Rather than gradient-based policy optimization, the agent learns in a training-free manner by accumulating and retrieving episodic memory (Appendix D.4) across training episodes. For each training episode the agent attempts navigation; the full action–observation trajectory and outcome are stored in the memory module at three granularities (step, trajectory, task). During test-time evaluation, the feedback signal is removed and the agent relies solely on retrieved memory to condition its policy.

Cross-benchmark evaluation. Beyond held-out SIMWORLD STUDIO test environments, all agents are evaluated on SimWorld-MMNav [102], an independently constructed navigation benchmark spanning three difficulty tiers (easy, medium, hard) in unseen UE5 environments. This evaluation directly measures whether navigation knowledge acquired in SIMWORLD STUDIO-generated worlds transfers beyond the training distribution.

D.3 Metric Definitions

We report four complementary metrics covering task success, path efficiency, and trajectory fidelity.

Notation. Let N denote the number of evaluation episodes. For episode i, let $P_i = (p_1, \ldots, p_T)$ be the executed trajectory and P_i^* the geodesic shortest-path reference, with lengths L_i and $L_i^* = d_i^0$ respectively. Let d_i be the geodesic distance from the agent's final position to the goal and $\delta = 1.0\,\text{m}$ the success threshold.

Success Rate (SR).

$$\mathrm{SR} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(d_i < \delta). \tag{1}$$

Binary indicator of task completion. The primary metric reported in the main paper.

Success weighted by Path Length (SPL).

$$\mathrm{SPL} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(d_i < \delta) \cdot \frac{L_i^*}{\max(L_i,\, L_i^*)}. \tag{2}$$

Penalizes inefficient paths; a successful agent that takes a much longer route than necessary scores below 1.

SoftSPL.

$$\mathrm{SoftSPL} = \frac{1}{N}\sum_{i=1}^{N} \max\!\left(0,\, 1 - \frac{d_i}{d_i^0}\right) \cdot \frac{L_i^*}{\max(L_i,\, L_i^*)}. \tag{3}$$

A continuous relaxation of SPL that rewards partial progress toward the goal rather than requiring binary success.

Normalized Dynamic Time Warping (nDTW).

$$\mathrm{nDTW} = \frac{1}{N}\sum_{i=1}^{N} \exp\!\left(-\frac{\mathrm{DTW}(P_i,\, P_i^*)}{\eta \cdot |P_i^*|}\right), \tag{4}$$

where $\mathrm{DTW}(\cdot,\cdot)$ is the dynamic time warping alignment cost, $|P_i^*|$ the number of waypoints in the reference path, and η a normalization constant ($\eta = 5$ in our experiments). nDTW measures trajectory fidelity beyond endpoint success, rewarding agents that closely follow the reference path.

D.4 Hierarchical Memory Design

The memory module extends the Generative Agents [50] and ExpeL [99] frameworks with *multi-level* updates that distill experience at step, trajectory, and task granularities, allowing strategies to be retrieved at the appropriate scope during inference.

Three-level memory structure.

- **Step-level memory (L1).** Stores fine-grained action–observation pairs within a single episode. Primarily used for within-episode self-correction (e.g., detecting a position revisit and triggering a recovery maneuver).
- **Trajectory-level memory (L2).** After each episode, the full trajectory is summarized into a compact natural-language record capturing navigation strategy, key decision points, and outcome. L2 records are indexed by environment features (obstacle density, path length tier) to enable retrieval of relevant past trajectories in new episodes.
- **Task-level memory (L3).** Periodically distills patterns across multiple L2 records into abstract navigation principles (e.g., "in dense urban environments, maintain a 45° offset from the goal bearing to avoid building corners"). L3 knowledge generalizes across environments and is injected into the system prompt as a high-priority context prefix.

Memory update. After each training episode, the module performs three sequential updates:

1. **L1 flush.** Step-level records from the completed episode are compressed and stored.
2. **L2 summarization.** An LLM summarizer condenses the episode trajectory and outcome into an L2 record, tagging it with environment metadata.
3. **L3 distillation.** Every 10 episodes, an LLM distiller identifies recurring patterns across recent L2 records and appends new L3 principles, replacing outdated ones.

Memory retrieval. At the start of each inference episode, the agent retrieves the top-k most relevant L2 records (by environment similarity) and the current L3 knowledge base. These are formatted into a structured memory block injected into the agent's context before the first action, giving the agent a head start grounded in prior experience.

D.5 Observation Modality Ablation

Table 11 reports a separate observation-modality study using a 100-episode training set and a 30-episode unseen test set as specified in the experiment record. It is not the 329-episode main evaluation, so its absolute SR values should not be compared directly with Table 2. No modality dominates across all model/split combinations: Qwen3.5-27B is strongest with depth on seen scenes and RGB on unseen scenes, while Qwen3.5-9B has its highest unseen SR with RGB-D. Text-only observations remain competitive in some settings. The reported deviations describe outcome variability and do not establish uncertainty across independent training seeds.

Table 11: **Ablation on observation modalities across model scales.** Each Qwen3.5 model is evaluated on ObjectNav with three vision configurations (RGB, Depth, RGB+Depth) and a text-only baseline where the agent receives goal distance and bearing scalars but no image. Metrics are success rate (SR) and SoftSPL (SS), both in $[0, 1]$ (↑). Best values within each (model scale, split, metric) group are highlighted in **bold**; tied values share the bold marker; tied zeros are not bolded.

Observation	Qwen3.5-27B				Qwen3.5-9B			
	Seen		*Unseen*		*Seen*		*Unseen*	
	SR↑	**SS↑**	**SR↑**	**SS↑**	**SR↑**	**SS↑**	**SR↑**	**SS↑**
RGB	.636±.487	.775±.227	**.833**±.408	**.721**±.373	**.136**±.347	**.279**±.332	0	.070±.113
Depth	**.659**±.479	**.784**±.198	.667±.516	.694±.356	.045±.211	.194±.308	0	.179±.255
RGB + Depth	.614±.493	.696±.308	.667±.516	.614±.386	.068±.255	.202±.290	**.167**±.408	**.223**±.331
Text only	.455±.504	.693±.267	.500±.548	.619±.369	.068±.255	.235±.276	0	.154±.162

E Case Study 3: Experimental Details

E.1 Structured Prompt Evolution

The embodied agent in Case Study 3 optimizes its navigation policy via a structured variant of GEPA [1]. Standard GEPA rewrites the entire system prompt at each update step, which risks overwriting effective strategies already in place. Our structured variant instead maintains an ordered list of prioritized decision rules $\mathcal{R}_t = [r_1, r_2, \ldots, r_k]$ that are injected verbatim into the system prompt in fixed priority order.

Rule format. Each rule r_i is a concise, actionable navigation heuristic expressed in natural language. Representative examples:

- *"If the bearing to the goal is within $\pm 30^\circ$ and no obstacle is visible, prefer* `move_forward`*."*
- *"If the same grid cell has been visited three or more times in the last ten steps, execute two consecutive* `turn_right` *actions to escape the loop."*
- *"When the reported distance to goal drops below 1.5 m, issue* `stop` *immediately."*

Rules are ordered by precedence: more specific rules (low-level reactive behaviors) appear first and override more general rules when both apply. A later rule takes effect only when no earlier rule fires.

Update mechanism. After each training episode, failure modes $\mathcal{F}_t$ are extracted from the trajectory and passed to the rule-synthesis LLM, which generates new rules $\Delta\mathcal{R}(\mathcal{F}_t)$ that address observed failures without contradicting existing rules:

$$\mathcal{R}_{t+1} = \mathcal{R}_t \ \cup \ \Delta\mathcal{R}(\mathcal{F}_t). \tag{5}$$

New rules are appended at the end of the list. A pruning step removes rules that have not been activated in the preceding 20 episodes. The list is capped at 30 active rules to prevent prompt bloat.

Failure mode taxonomy. The rule-synthesis prompt presents the full action–observation trajectory alongside a structured failure taxonomy:

1. **Directional errors.** Turning when a clear forward path exists; misalignment between bearing and action.
2. **Loop behavior.** Revisiting the same position repeatedly; oscillating between two nearby cells.

3. **Goal proximity failure.** Stopping too far from the goal; overshooting and failing to recover.
4. **Obstacle avoidance failure.** Repeated collisions with the same obstacle type (e.g., building corner).
5. **Timeout.** Reaching the 500-step budget without issuing `stop`.

For each detected failure, the LLM synthesizes one to three targeted rules. The synthesis prompt also receives the current rule list $\mathcal{R}_t$ to enforce non-contradiction.

E.2 Difficulty Curriculum Parameterization

SIMCODER controls episode difficulty via three axes: geodesic path length, initial heading offset (angular deviation from the goal direction at episode start), and dynamic obstacle density (fraction of navigable area occupied by movable obstacles).
Table 12 lists the eight difficulty levels (L0–L7) with their exact parameter ranges. Levels are designed so that each axis increases monotonically: longer paths demand extended planning, larger heading offsets require deliberate reorientation before progress can begin, and higher obstacle density increases path variability and collision risk.

Table 12: **Difficulty level parameterization.** Path length is the geodesic distance from start to goal; heading offset is the initial angular deviation from the goal direction; obstacle density is the fraction of navigable area occupied by dynamic obstacles.

Level	Path Length (cm)	Heading Offset (°)	Obstacle Density
L0	400– 800	0– 15	0.00
L1	600– 1000	0– 30	0.05
L2	800– 1400	0– 45	0.10
L3	1000– 1800	0– 60	0.15
L4	1200– 2200	0– 90	0.20
L5	1500– 2600	0– 120	0.25
L6	2000– 3000	0– 150	0.30
L7	2500– 3500	0– 180	0.35

Mastery-gating thresholds. SIMCODER advances to the next level only when the agent's rolling success rate $\bar{S}_t$—averaged over the most recent 5 epochs—exceeds a level-specific threshold τ_ℓ:

$$\ell_{t+1} = \min\big(7,\ \ell_t + \mathbf{1}\big[\bar{S}_t \geq \tau_{\ell_t}\big]\big)\,. \tag{6}$$

Thresholds decrease with difficulty to maintain challenge as tasks grow harder: $(\tau_0, \ldots, \tau_7) = (0.80,\ 0.75,\ 0.70,\ 0.65,\ 0.60,\ 0.55,\ 0.50,\ 0.45)$. This mastery-gating scheme mirrors the zone-of-proximal-development principle [69]: the agent is exposed to the next challenge only once it has demonstrated reliable competence at the current level.

E.3 Training and Evaluation Protocol

Co-evolution loop. Each training epoch proceeds in four steps:

1. **Episode generation.** SIMCODER generates 20 navigation episodes at the current difficulty level ℓ_t using the SIMWORLD STUDIO embodied interface.
2. **Agent rollout.** The embodied agent attempts all 20 episodes; action–observation trajectories and outcomes are recorded.
3. **Policy update.** The prompt-evolution module updates $\mathcal{R}_t$ from extracted failure modes.
4. **Difficulty advance.** The rolling success rate $\bar{S}_t$ is computed; if $\bar{S}_t \geq \tau_{\ell_t}$, SIMCODER advances to ℓ_{t+1}.

The benchmark evaluation on SimWorld-MMNav is run every 5 epochs. Training runs for 25 epochs total (500 episodes).

Baseline conditions.

- **Co-evolving environment + learning.** The full co-evolution loop as described above. SIMCODER adapts difficulty; the agent accumulates rules.

- **Fixed-difficulty environment + learning.** The same prompt-evolution mechanism, but SIM-CODER holds difficulty fixed at L3 (medium) throughout. Isolates the contribution of the adaptive curriculum from the rule-learning mechanism.
- **Random-difficulty environment + learning.** The agent accumulates rules while the difficulty level is sampled from L0–L7 rather than selected by mastery.
- **No learning.** A fixed system prompt with no rule accumulation and no curriculum adaptation. Serves as the zero-shot LLM baseline.

All conditions use Qwen3.5-9B and the same initial system prompt with an empty rule list.

Evaluation. Final performance is measured on SimWorld-MMNav [102], a held-out benchmark of 329 episodes spanning easy, medium, and hard difficulty tiers in independently authored UE5 environments unseen during training. We report Success Rate (SR) as the primary metric.

F Additional Evaluation and Operational Analysis

This section reports additional experiments on generation scale, evaluation consistency, operational cost, generator comparisons, policy adaptation, and manipulation.

F.1 Expanded Scene-Generation Evaluation

We generated 200 scenes per builder under the scene-generation evaluation setting. PF is prompt fidelity, SRF spatial-relation fidelity, and LAES layout aesthetics. Scores are normalized to $[0, 1]$. The mean column averages the three rubric dimensions; SD is across scenes, not training seeds. This extends the original nine-scene diagnostic to 200 scenes per builder.

Table 13: Expanded generation evaluation: mean $\pm$ scene-level SD; 200 scenes per builder.

Builder	PF	SRF	LAES	Mean
Qwen3.5-27B	0.493 ± 0.041	0.560 ± 0.038	0.445 ± 0.029	0.499 ± 0.031
Qwen3.5-2B	0.084 ± 0.022	0.123 ± 0.027	0.147 ± 0.024	0.118 ± 0.021

F.2 Cross-Judge Agreement

We re-evaluated the same scenes with three independent VLM families: Claude Opus 5, GPT-5.6 sol, and Kimi K3. PF measures requested semantic content, SRF correct spatial relations, and LAES layout coherence and aesthetics. Table 14 gives dimension means and SD across per-scene mean scores. Pearson/Spearman coefficients and mean absolute difference (MAD) are computed over the per-scene mean of PF, SRF, and LAES.

All pairwise correlations exceed 0.80, supporting consistent scene-quality rankings across judge families despite calibration differences. Mean judge scores span 0.398–0.481, and mean absolute differences span 0.097–0.121.

Table 14: Cross-judge scores, normalized to $[0, 1]$.

Judge	PF	SRF	LAES	Mean	SD
Claude Opus 5	0.423	0.373	0.397	0.398	0.072
GPT-5.6 sol	0.507	0.507	0.411	0.475	0.089
Kimi K3	0.473	0.477	0.493	0.481	0.106

Table 15: Agreement over per-scene mean rubric scores.

Judge pair	Pearson r	Spearman ρ	MAD
Opus 5 / GPT-5.6 sol	0.828	0.816	0.102
Opus 5 / Kimi K3	0.821	0.805	0.121
GPT-5.6 sol / Kimi K3	0.830	0.836	0.097

F.3 Generation Runtime, API Cost, and Scaling Limits

Two complete town-generation traces include verification and repair (Table 16). They contain 1,633/2,205 actors over approximately 59,000 m^2, at about \$0.0024/\$0.0022 per actor. These measurements cover two complete town-generation runs. Generation is offline; saved environments can be reused without regenerating scenes.

Table 16: Measured end-to-end traces. Verifier is the logged internal score, not cross-judge scene quality.

Scene	Minutes	API cost	Turns	Verifier /100	Actors	Building types
Desert town	5.5	\$3.94	101	99.8	1,633	64
Winter town	8.9	\$4.77	158	100.0	2,205	125

Table 17: Stage-level ranges for the two traces. Model-API cost excludes local compute.

Stage	Minutes	API cost
Initial planning and code generation	1.7–2.5	\$2.20–\$2.55
UE tool execution	1.3–2.0	\$0
Rule-based physical checks	0.4–0.6	\$0
VLM verification and agent repair	1.6–3.1	\$1.74–\$2.22
Task derivation and Gym export	0.5–0.7	\$0
Total	5.5–8.9	\$3.94–\$4.77

The reported verification summary is 2.7 rounds and 62 corrective tool actions per scene. Local engine execution, checks, and export incur computation even with zero API charge. Hardware time, energy, and engineering effort are not priced in these charges.

Throughput projection. At 5–10 minutes per scene, 20 continuously active workers with ideal linear scaling would produce $20 \times 1440/(5\text{–}10)$ = 2,880–5,760 scenes/day, or 10,000 scenes in approximately 1.7–3.5 days. This is a projection, not a measured deployment. UE workers, model/API concurrency, failures, skill-registry coordination, and shared-resource contention can reduce throughput. The measured scale is 200 scenes per builder.

Asset coverage. The registry exposes 17,955 assets; the built-in building-placement skill covers 127 building classes. Registry size is a vocabulary ceiling, not per-scene utilization or a diversity metric. Licensed assets can be registered, external retrieval exposed through MCP/Python, and generated meshes imported with scale, collision, material, and physics checks. Asset sources are decoupled from the generator, so new assets can use the existing composition and verification pipeline.

F.4 Comparison with Scalable Generators

We compare 100 scenes per generator using the same PointNav policy, training budget, and held-out evaluation protocol. ProcTHOR uses procedural room specifications, Holodeck language-conditioned construction, and SIMWORLD STUDIO engine-level code with reusable skills.

Table 18: Comparison with ProcTHOR and Holodeck using 100 scenes per generator. All metrics are higher-is-better.

Generator	Layout diversity	Physical validity	PointNav transfer SR (%)
ProcTHOR	0.58	0.97	68
Holodeck	0.64	0.86	66
SIMWORLD STUDIO	0.79	0.99	74

SIMWORLD STUDIO obtains layout diversity of 0.79, compared with 0.58 for ProcTHOR and 0.64 for Holodeck, while retaining physical validity of 0.99. PointNav transfer SR is 74%, compared with 68% and 66%, respectively. ProcTHOR and Holodeck primarily cover indoor scenes, whereas SIMWORLD STUDIO also supports outdoor environments. Registry size is a separate measure: the ProcTHOR released 10,000-house corpus contains 996 observed assets, while 17,955 registered assets are available to SIMCODER. These counts compare observed/available vocabulary ceilings rather than per-scene utilization.

F.5 Memory and Trajectory-SFT Control

Memory keeps weights frozen and retrieves rules distilled from successful and failed trajectories. It supports local or API policies without a separate weight-training job. The trajectory-SFT control instead updates Qwen3.5-9B on filtered target trajectories collected in generated environments.

Table 19: Qwen3.5-9B adaptation comparison; SR values are percentages.

Adaptation	Weight update	PointNav SR	ObjectNav SR
None	No	0.40	0.00
Trajectory SFT	Yes	6.80	12.90
Memory	No	7.43	14.62

Memory improves over the unadapted policy by 7.03 PointNav and 14.62 ObjectNav percentage points, while avoiding weight updates. It slightly outperforms the trajectory-SFT control in this experiment. Both methods remain compatible with the generated environments.

F.6 Interpreting Co-Evolution Feedback

Trajectory failures update embodied-policy rules, whereas aggregate mastery updates environment difficulty. Repeated cell visits can yield the rule "if the same cell has been visited at least three times in the last ten steps, execute two consecutive right turns." Building-corner collisions can yield a task-level rule to maintain a 45° offset from the goal bearing in dense urban scenes. These are agent-side corrections, not direct instructions to place an obstacle in the next scene.

When rolling success crosses the mastery threshold, the L3-to-L4 transition increases path length from 1,000–1,800 cm to 1,200–2,200 cm, heading offset from 0–60° to 0–90°, and obstacle density from 0.15 to 0.20 (Table 12). Otherwise the level is retained. Figure 7 uses the same training-episode count and independently authored 329-episode benchmark across conditions. Fixed L3 without rules is a separate control, although its SR equals the unadapted baseline (50%). Test outcomes never enter curriculum gating, rule updates, or generation.

F.7 Manipulation Protocol and Scope

The MuJoCo adapter retains specification, construction, checking, repair, and Gym evaluation while replacing backend execution. Pick-and-place layouts span 0/1/2 obstacles and are checked for reachability, transport distance, gripper clearance, and path blocking. In the 100-task feasibility study, both a scripted expert and Qwen3.5-27B with images and object poses solve 100/100.

The learning study uses 50 training and 50 disjoint test scenes. Both Qwen3.5-27B conditions omit the hand-written strategy block; only the adapted condition uses distilled training experience (Table 3). The observed five-task improvement corresponds to a 10-percentage-point success-rate gain. Median final placement distance decreases from 0.084 m to 0.040 m; mean steps decrease from 8.2 to 5.4, and mean collisions from 87.7 to 52.3.

This study evaluates a single tabletop manipulation family in simulation. Long-horizon, mobile-manipulation, multi-agent, and real-robot evaluations remain future work.

G Prompt Examples

Text-to-Scene (Hard). *"Design a full residential neighborhood. Build two parallel streets with buildings on both sides: place 3 buildings along the north side and 3 buildings along the south side of each street (total 6 buildings in two rows). Connect the streets with a cross road. Line all roads with trees on both sides. Create a central park between the two streets with tables, couches, and trash bins. Add a fire hydrant at every street intersection. Place scooters and carts parked along the roads. Mark one intersection with road cones and road blockers as a construction zone."*

Image+Text-to-Scene (Hard). *"Build a dense urban block matching this aerial photo: multiple buildings arranged in a grid pattern with streets between them, trees lining the streets, and vehicles and street furniture throughout."* (Accompanied by a real aerial photograph of a city block.)

Scene Editing (Hard). *"Expand the plaza into a larger district: add 2 new buildings on the north side. Add roads connecting the new buildings to the existing ones. Plant 6 more trees to line the new roads. Create a marketplace area with 3 tables and 2 carts near the center. Add road cones and road blockers to mark a construction zone near the new buildings. Place fire hydrants at each road intersection and trash bins along the sidewalks."*

H Qualitative Examples

The qualitative examples in this appendix use the model snapshots available at the time of figure generation. They are provided only as visual examples and are not included in the quantitative comparisons in Table 1.

H.1 Text-to-Scene

Prompt P1

Build a downtown city-block scene at a 4-way cross intersection in Unreal Engine 5, set during the morning. The aesthetic, building heights, density, and how lively each side feels are completely your call - make it feel like a real downtown intersection.

It needs to be a proper 4-way cross, not a T-intersection, with one road running along the X axis and another cutting through it along Y. Make sure the seam where they meet is clean. Nothing should overlap the road or sidewalk - buildings stay outside the road footprint, cars stay in their lanes, pedestrians stay on sidewalks or in crosswalks. Run everything in a single script call, then take one screenshot, then stop.

Everything else is yours. How long each road is, which quadrants get tall glass towers versus shorter brick mid-rises, whether buildings are packed tight or have occasional plazas and setbacks - that's your composition. Same goes for sidewalk dressing like trees, café tables, and benches, how many people are out, car patterns, and whether the two roads feel like the same neighborhood or two different ones. Try to get specific and make real decisions rather than just defaulting to whatever looks average - the more deliberate the choices, the more the scene has a point of view.

The skill library is staged and ready to use. Before writing the script, briefly sketch out what kind of downtown this is, what the focal points are, and how the density and lighting support the mood - then write it.



Claude Opus 4.7 | Qwen 3.5-27B | Qwen 3.5-9B

Figure 9: **Qualitative Example P1.** Output scenes generated by three model backbones given the same downtown city-block intersection prompt.

Prompt P2

Build a dense urban avenue scene in Unreal Engine 5, set during the morning. It's a single straight street running east–west, lined with buildings on both sides. The aesthetic, the density, the mix of building heights, and how lively it all feels are completely your call - make it feel like a real neighborhood you'd actually recognize.

It needs to be a single straight road, no intersections or curves. Nothing overlaps the road or sidewalk - buildings stay outside the road footprint, cars stay in their lanes, pedestrians stay on sidewalks or in crosswalks. Run everything in a single script call, then take one screenshot, then stop.

Everything else is yours to decide. How long the street is, whether both sides get the same treatment or feel like different neighborhoods, whether buildings are packed wall-to-wall or have the occasional gap, plaza, or setback. Think about the sidewalk too - is it sparse and clean, or are there trees everywhere, café tables spilling out, benches? How many people are out, how are the cars behaving, does the street feel chaotic or orderly? The more deliberate those choices are, the more the scene has its own personality.

The skill library is staged and ready. Before writing the script, briefly think through what kind of street this is, what the focal points are, and how the density and lighting support the mood - then write it.



Claude Opus 4.7 | Qwen 3.5-27B | Qwen 3.5-9B

Figure 10: **Qualitative Example P2.** Output scenes generated by three model backbones given the same dense urban avenue prompt.

H.2 Image+Text-to-Scene

Prompt P1

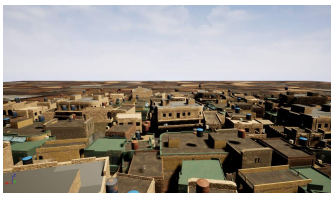
Create a large Middle Eastern desert village that matches the attached reference image. The reference shows a dense settlement of low-rise buildings spread across desert terrain - your job is to recreate that feeling using the tools available.

A few things are fixed: start by establishing the desert ground, then build the village with somewhere between 150 and 200 houses, then set up the lighting. The village builder already handles clutter internally so don't add random props on top of that. Pick your own random seed.

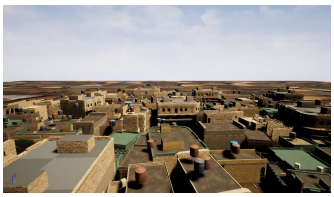
Everything else is a judgment call based on what you see in the reference. How many houses, how tightly packed they are, how organic or grid-like the layout feels - those are yours to decide. Think about whether rooftops should have water tanks and antennas, whether some houses get compound walls, and whether there's debris and clutter visible between buildings. Beyond the village itself, think about where dirt roads cut through, where fences sit, how many water tanks and market tents are scattered around the edges, and where they'd naturally fall. Study the reference and let it guide those decisions rather than just picking middle-of-the-road values for everything.



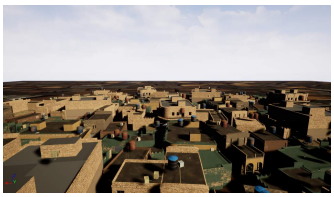
Reference Image



Claude Opus 4.7



Qwen 3.5-27B



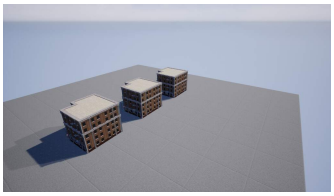
Qwen 3.5-9B

Figure 11: **Qualitative Example P1.** Top: prompt and reference image. Bottom: rendered UE5 screenshots from each model backbone.

H.3 Scene Editing

Prompt P1

Build a two-sided residential street in the current scene, which already has three starting buildings and three trees. Keep one existing building, remove the other two, then fill out both sides using exactly two building types, the kept one plus one other small building of your choosing. Check the scene first, then remove the unwanted buildings, lay the road, and build outward. The north side incorporates the kept building with additional buildings extending west and east; the south side gets its own row facing north. Once buildings are placed, add pedestrians on both sides, vehicles in the road, and finish with a cloudy sky. Street length between 18,000 and 28,000, road width 800–1,100, sidewalk width 400–700, road center Y between -2,500 and -1,700. North row: 3 to 6 buildings on each side of the existing house. South row: 8 to 14. Pedestrians: 40 to 70 per sidewalk. Vehicles: 10 to 16. Two building types only.



Original Scene



Claude Opus 4.7



Qwen 3.5-27B

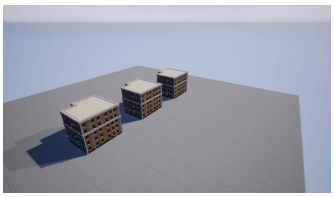


Qwen 3.5-9B

Figure 12: **Qualitative Example P1.** Top: editing prompt and the original scene prior to modification. Bottom: rendered UE5 screenshots showing each model's edited scene, built on top of the same starting configuration.

Prompt P2

Build a two-sided residential street in the current scene, which already has three starting buildings and three trees. Remove two of them, keep one, and build out both sides using a fixed pool of four small 2-to-3 storey buildings, mixed randomly per tile. Check the scene first, remove the unwanted buildings, lay the road, then build outward. North side incorporates the kept building with rows extending east and west; south side gets its own row facing north. Add pedestrians on both sides, vehicles in the road, and finish with a cloudy sky. Pick real values: street length 18,000–28,000, road width 800–1,100, sidewalk width 400–700, road center Y between -2,500 and -1,700. North row: 3 to 6 buildings each side. South row: 8 to 14. Pedestrians: 40 to 70 per sidewalk. Vehicles: 10 to 16.



Original Scene



Claude Opus 4.7



Qwen 3.5-27B



Qwen 3.5-9B

Figure 13: **Qualitative Example P2.** Top: editing prompt and the original scene prior to modification. Bottom: rendered UE5 screenshots showing each model's edited scene, built on top of the same starting configuration.

H.4 Iterative Scene Development

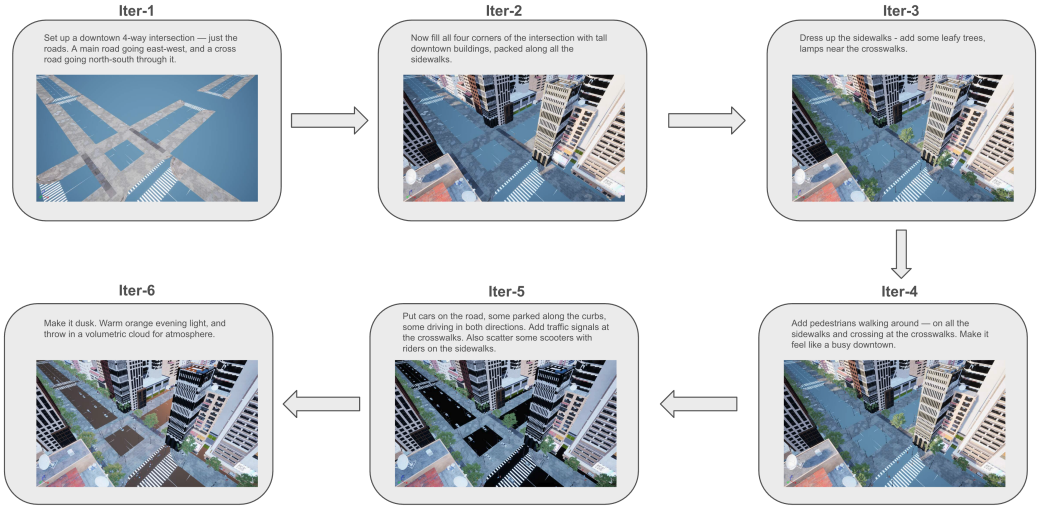


Figure 14: **Iterative scene development over six steps.** Starting from a bare 4-way road intersection (Iter-1), the scene is progressively enriched through a sequence of natural language editing instructions: tall downtown buildings are added at each corner (Iter-2), sidewalks are dressed with trees and lamps (Iter-3), pedestrians populate the crosswalks and sidewalks (Iter-4), cars, scooters, and traffic signals are introduced (Iter-5), and finally the lighting is shifted to dusk with warm orange tones and volumetric clouds (Iter-6). Each iteration modifies only what is specified without rebuilding the scene from scratch, demonstrating the platform's ability to support compositional, instruction-driven scene development.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations appear in Section 5 and Appendices A.7 and F.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: This work is not mainly a theoretical contribution.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [No]

   Justification: The paper describes the generation pipeline and navigation protocols. Additional experiments are reported as aggregate results; not all run-level configurations are provided.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The submission provided a review-time code repository: `https://anonymous.4open.science/r/SimWorldStudio-Submission-42CB`.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [No]

   Justification: The paper specifies the main case-study setups and additional experimental conditions, but does not enumerate every run-level hyperparameter for the added controls.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The expanded generation study reports scene-level SD and the judge study reports agreement coefficients. Most downstream and ablation results are point estimates without independent-seed confidence intervals; no statistical-significance claim is made.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [No]

Justification: Appendix A.6 gives system requirements; Appendix D reports the eight-H100 navigation setup and runtime. Appendix F.3 gives two measured generation traces. Per-experiment hardware and total compute are not exhaustively reported.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Yes. This paper follows Code of Ethics.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Yes we discussed in appendix A.8.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [Yes]

    Justification: We release the platform for research use with documented dependencies and license constraints for assets. The system does not release trained harmful policies or scraped personal data. Potential misuse of automated scene generation is discussed in Appendix A.8, and users are expected to comply with asset licenses and deployment policies.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Existing software, model APIs, and asset libraries are credited in Appendix A.9, including source/owner, version or access path, and license or terms of use.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The released repository documents the platform code, generated environment format, task construction pipeline, and reproduction instructions. The new assets introduced by this paper are code, prompts, task-generation scripts, and generated-environment metadata rather than human-subject data.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Paper does not involve crowdsourcing nor research with human subjects.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: Paper does not involve crowdsourcing nor research with human subjects.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: Yes, we describe all the usage of LLMs on our core components.