



SceneOrchestra: Efficient Agentic 3D Scene Synthesis via Full Tool-Call Trajectory Generation

Yun He · Kelin Yu · Matthias Zwicker

University of Maryland, College Park

Plan the whole tool-call trajectory up front — better scenes, in half the time



Task & Motivation

Generate realistic, controllable, and physically valid 3D scenes from natural language.

Rule-based, data-driven, and LLM-based methods are all monolithic: each one achieves physical validity, realism, or controllability, but never all three at once.

Agentic frameworks avoid that trade-off by combining specialised tools in a single workflow. The open question is how to orchestrate the calls to those tools.

Limitations of Agentic Workflows

01 Heuristic orchestration

In today's agentic workflows an off-the-shelf LLM picks each tool and its parameters from hand-written use cases, so the execution flow is often far from optimal: detours, redundant calls, weaker scenes.

02 Step-by-step review

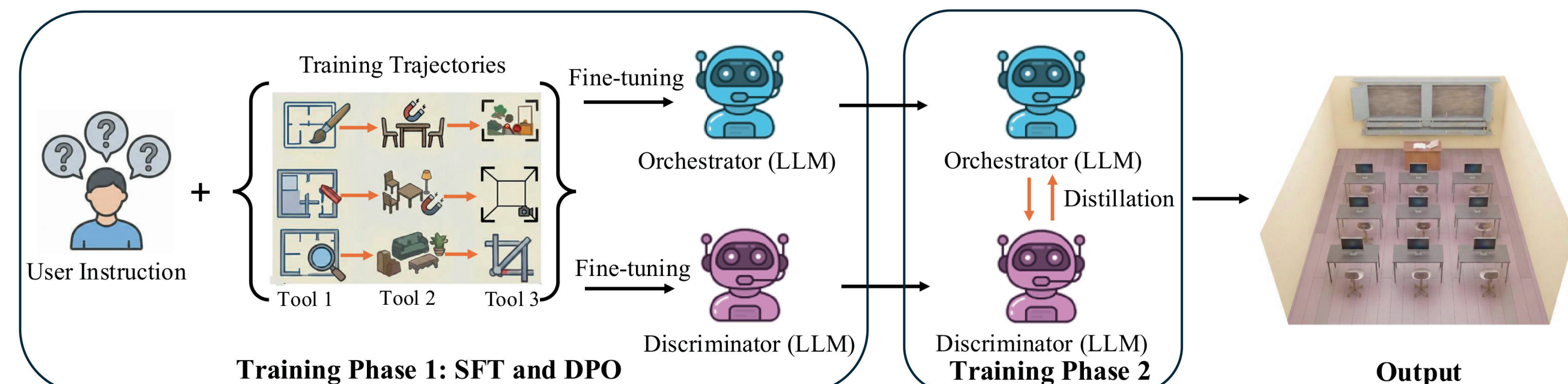
The same loop renders every intermediate scene and asks the LLM to review it before moving on. On a single A6000 that stretches one scene to 91 minutes.

The bottleneck is the orchestration policy, not the tools.

SceneOrchestra

Predict the full tool-call trajectory — then execute it once.

"Design a computer classroom with several individual workstations. Each workstation has a desktop monitor and a rolling chair, all facing the front of the room. Place a teacher's desk at the front center, with a large blackboard on the wall behind it."

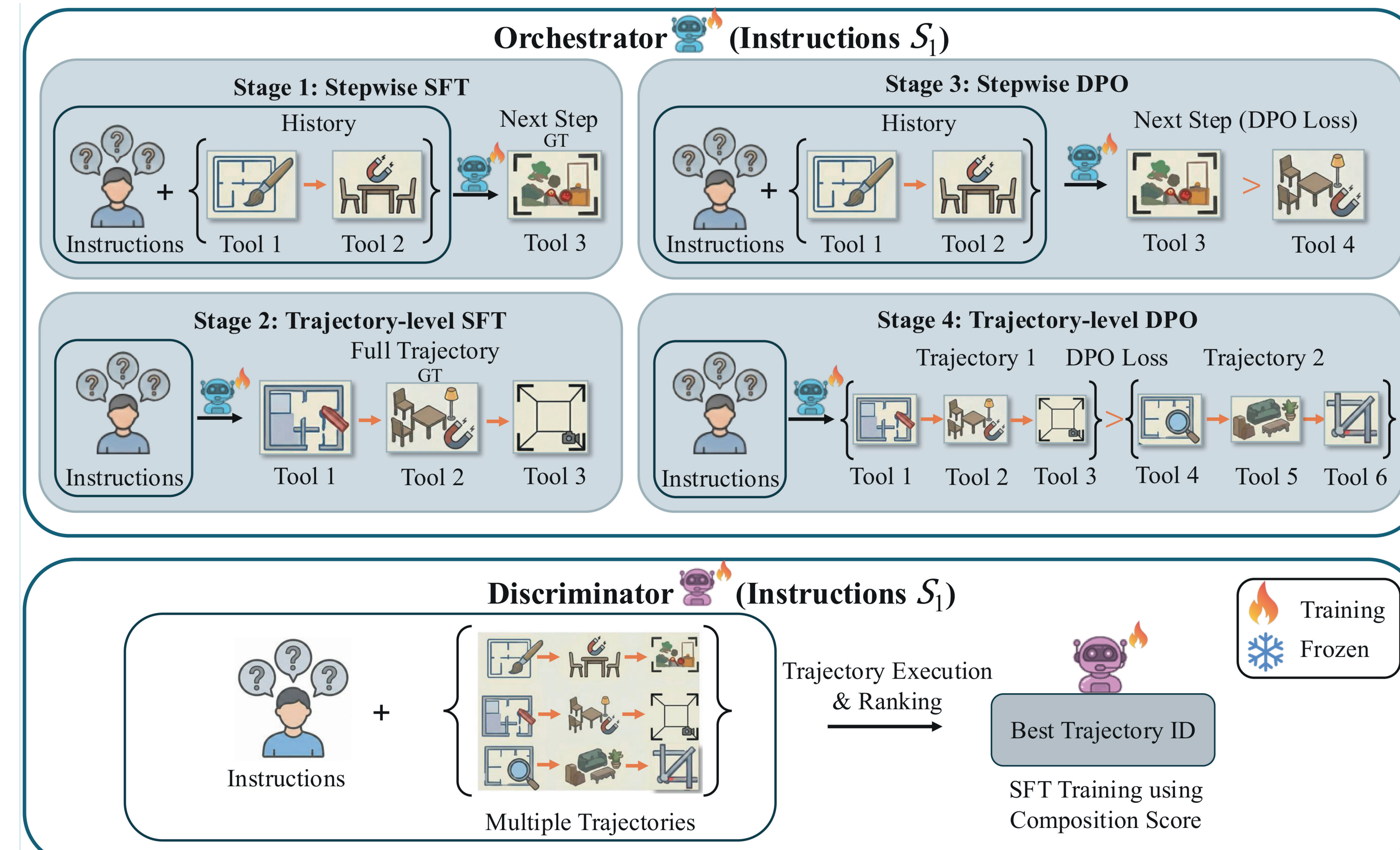


- The orchestrator turns a user instruction into a complete tool-call trajectory in a single pass.
- A discriminator ranks the candidate tool-call trajectories, and its ranking ability is then distilled from the discriminator back into the orchestrator.
- At inference the discriminator is dropped: the orchestrator alone predicts the trajectory, and the tools execute it end to end in one shot.

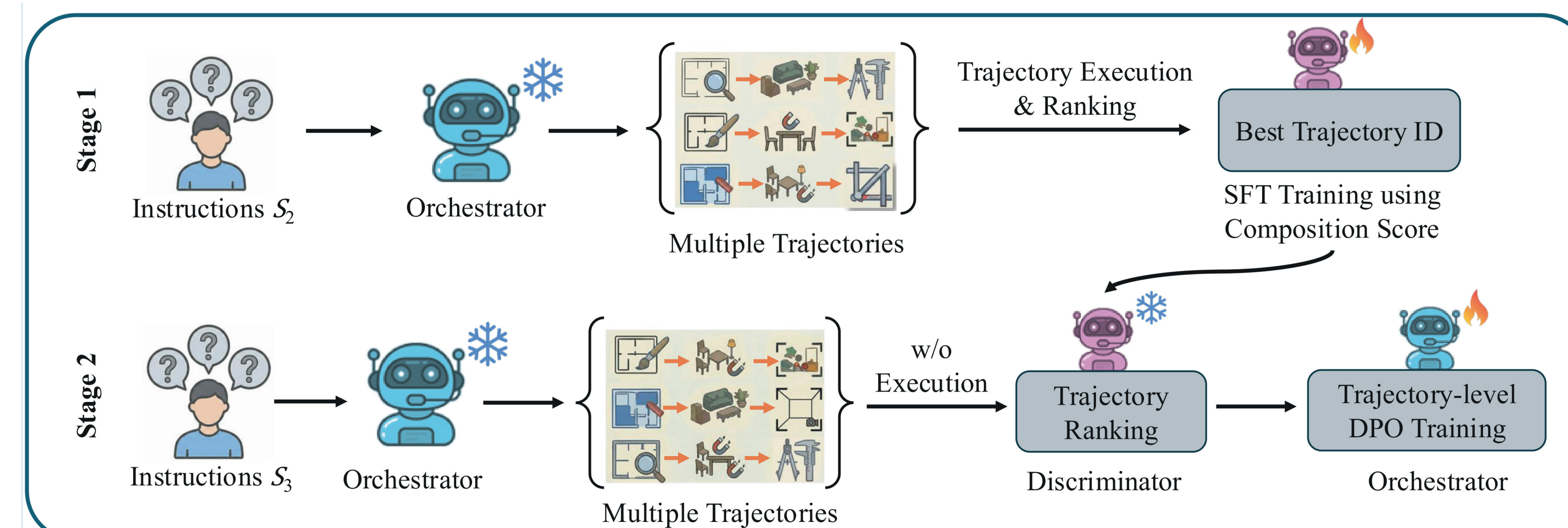
Method: Two-Phase Training

The orchestrator learns to plan complete tool-call trajectories; the discriminator learns to rank them, and distils that ability back into the orchestrator.

Phase 1 · Independent training



Phase 2 · Interleaved training



INFERENCE

At test time only the orchestrator runs. It emits the entire tool-call trajectory in a single pass, and the tools execute it end to end — no intermediate rendering, no review loop, no discriminator.

Results: Template Instructions



Average over 10 room types · 5 generations per instruction

Method	#Obj ↑	#OB ↓	#CN ↓	Real. ↑	Func. ↑	Lay. ↑	Comp. ↑	Time ↓
LayoutGPT	7.8	1.1	1.2	6.0	6.2	5.5	3.6	2.4
Holodeck	26.2	0.0	8.6	7.8	7.8	6.4	5.8	12.4
I-Design	12.3	0.3	0.1	6.9	7.6	6.3	4.5	6.1
SceneWeaver	27.0	0.0	0.2	8.8	8.7	7.2	7.5	91.0
Ours	31.1	0.0	0.0	9.1	9.1	7.9	8.1	39.7

HUMAN STUDY

18 participants · pairwise preference and diversity

Metric	vs LayoutGPT	vs Holodeck	vs I-Design	vs SceneWeaver
Preference	100%	84.4%	95.6%	76.7%
Diversity	100%	88.9%	96.6%	80.8%

Across every baseline, participants preferred SceneOrchestra for both quality and diversity.

Complex Instructions: Fine-Grained Control

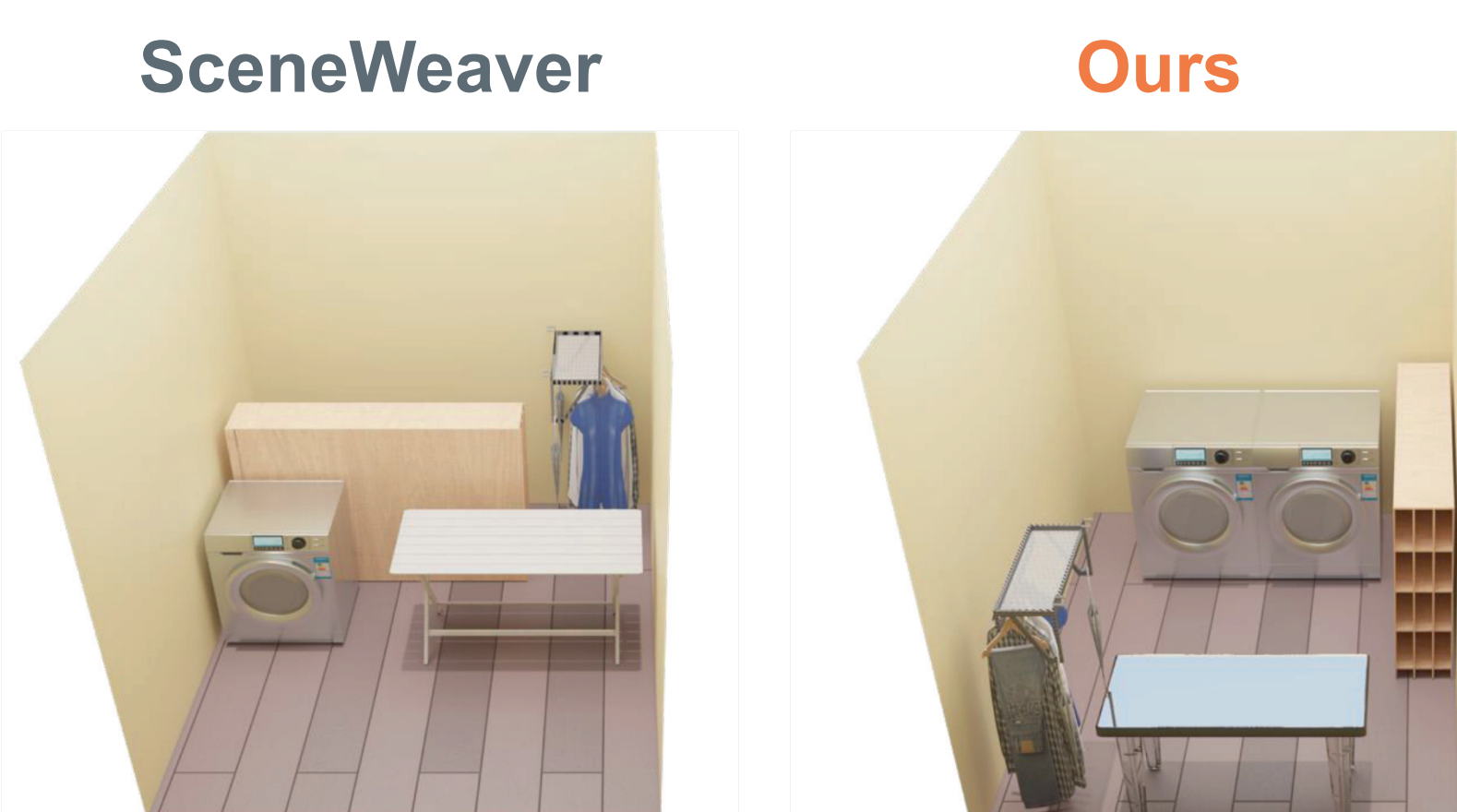
1. Garage

"Design a garage with a car parked at the center, two bicycles leaning against the back wall, and a workbench along the side wall."



2. Laundry room

"Create a laundry room with two front-loading washing machines side by side against the back wall, a countertop with a folding area directly opposite, and a drying rack with hanging garments in the corner."



3. Compact bathroom

"Create a compact bathroom with a shower stall in the corner. Add a toilet, a sink with a small vanity and towel rack, and a bath mat on the floor."



4. Home office

"Design a home office with a desk and a computer monitor as the central workspace, paired with a chair. Add shelving units with books and small decorative items for storage and decoration."



5. Nursery room

"Generate a nursery room with a baby bed and a small changing table for infant care. Add a rocking chair and a storage cabinet stocked with toys and baby supplies."

