Coding Agents for Generalized Task and Motion Planning Problems

Matteo Merler[2], Bowen Li[3], Josh Roy[1], Yichao Liang[4], Qianwei Wang[1], Yixuan Huang[1], and Tom Silver[1]

Website · Code

***Abstract*— Task and motion planning (TAMP) problems remain difficult even with full observability and object-centric states because discrete decisions are tightly coupled to geometric, kinematic, and dynamic constraints. Generalized TAMP addresses this difficulty by exploiting regularities across problem instances to reduce planning effort on new instances. However, existing methods require substantial TAMP-specific engineering. We investigate whether coding agents can automate this process by synthesizing programs that generalize across instances. Given a task description and simulator access, each agent chooses how to interact with the environment while developing a program within a fixed synthesis budget. The program is then frozen and evaluated on unseen instances. We evaluate *Claude Code* (Opus 5) and *Codex* (GPT-5.6 Sol and GPT-6 Astra) on 28 simulated environments from KinDER and PDDLStream, with object counts beyond those evaluated in the original benchmark. Across all program synthesis methods, we evaluate 980 generated programs on 100 held-out instances each, 98,000 evaluation episodes in total. Overall, we find that coding agents are surprisingly effective at generalized TAMP: all three agent configurations outperform hand-engineered planners, one-shot generation, and an LLM-based generalized planning baseline in mean success (56% to 95% versus 47% for the planners, on the 16 environments where a planner is available). As object counts grow, the agents' programs maintain higher success than the planner, using an order of magnitude less computation per instance on average. Logs show agents using interaction to calibrate physical models, test edge cases, and refine strategies. We release all code, including the full prompts given to the agents. These findings suggest that coding agents are a strong baseline for generalized TAMP.**

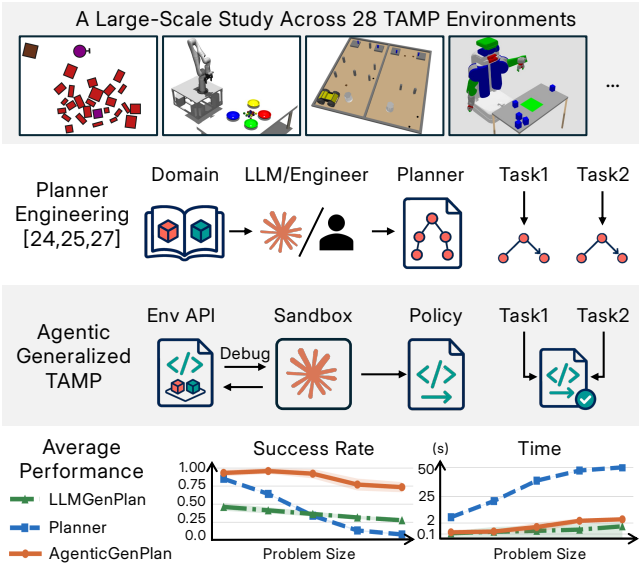

Fig. 1: **Coding agents for generalized TAMP.** A coding agent (*Claude Code* with Opus 5) synthesizes a programmatic policy shared across problem instances, without hand-designed planning components. Compared with the planner and LLM-based generalized planner, the agentic method achieves a higher success rate while maintaining efficiency. Bottom: success rate and computation time per instance as the number of objects grows, averaged over environments with a planner available and multiple object counts.

I. Introduction

We are interested in the extent to which state-of-the-art coding agents can solve the constrained manipulation problems that typify task and motion planning (TAMP). Even with full observability and object-centric states, these problems remain formally hard [1] and challenging in practice because horizons are long, feedback is sparse, and geometric, kinematic, and dynamic constraints are tightly coupled [2]–[7]. To mitigate these difficulties, generalized TAMP [8]–[14] exploits regularities across problem instances to produce reusable solutions, for example by learning samplers, feasibility predictors, search heuristics, or abstractions. We ask whether coding agents can similarly discover regularities that enable fast and effective planning, while relying on far less TAMP-specific scaffolding than previous methods.

Prior evidence points in opposite directions. Coding agents are improving rapidly on software-engineering benchmarks [15], [16], and large language models (LLMs) have shown increasing success in classical planning domains [17], [18]. Yet when LLMs are asked to make the geometric and physical decisions within a TAMP system, they perform poorly, even when the relevant geometry is included in the prompt [19], [20]. It therefore remains unclear whether advances in coding transfer to the physical reasoning that distinguishes TAMP. Either answer matters for the field. Success would suggest that coding agents can reduce much of the domain-specific engineering required by TAMP methods; failure would help identify concrete limitations of current agents.

To answer this question, we present a large-scale, systematic study of off-the-shelf coding agents for generalized TAMP (AgenticGenPlan). The study covers 28 environments, seven program synthesis methods, five runs per method per environment, and 100 held-out test instances per resulting program, 98,000 evaluation episodes in total. For each environment, we give a coding agent a task description, simulator access, and a fixed synthesis budget (Figure 1). Within this budget, the agent chooses which experiments to run, writes code to probe the simulator, and uses the results to develop and test its program. The agent returns a single program,

[1]Princeton University, [2]Fondazione Bruno Kessler, [3]Carnegie Mellon University, [4]University of Cambridge
tsilver@princeton.edu

which is frozen and evaluated on unseen instances, with no LLM involved at test time. In our main setting, the agent receives no environment source code and no hand-designed planning abstractions. Beyond the task description and simulator access, nothing is engineered for the agent.

We evaluate this approach with two different agentic backends (*Claude Code* with Opus 5 and *Codex* with GPT-5.6 Sol and GPT-6 Astra) on environments spanning kinematic and dynamic tasks in two and three dimensions, including the KinDER benchmark [20] (comprising 25 environments across four families) and three PDDLStream domains [19]. We compare the synthesized programs with TAMP planners given the hand-designed predicates, operators, samplers, and skills supplied by the benchmarks, and we measure both success and runtime as the number of objects increases. We also compare interactive synthesis with the LLM-based generalized planning method LLMGenPlan [21] and the one-shot generation variant, both given environment source code. We additionally give *Claude Code* and *Codex* with GPT-6 Astra the environment source code. This setting serves as a reference for how the agents perform with complete knowledge of the environment. We release all code, including the full agent prompts.

Overall, all three agent configurations outperform the hand-engineered planners on the 16 environments where one is available: *Claude Code* averages 82% success, 1.7 times the planner's 47%, and *Codex* averages 95% with GPT-6 Astra and 56% with GPT-5.6 Sol. As object counts grow, they maintain higher success and low computation times, while the planner takes longer and solves fewer instances (Figure 1). All three agents also outperform one-shot generation and LLMGenPlan in mean success over all 28 environments, and source access enables successful programs in environments where main-setting runs fail.

Analysis of the synthesized programs and interaction logs shows agents using targeted probing to infer environment dynamics and developing strategies that exploit regularities across instances. Some programs narrow a search over actions by first identifying which obstacles block a route to the goal, then planning how to move them; others adapt a fixed manipulation sequence to each instance without searching. The agents also discovered unexpected strategies, including non-prehensile maneuvers and uses of the environment layout that, to our knowledge, have not appeared in prior work on these benchmarks (Figure 2). We consider this qualitative evidence among the strongest in the study, both for what it says about agentic physical reasoning and because strategies absent from any published solution are hard to attribute to memorized training data. Failures remain: some dynamic three-dimensional environments are still largely unsolved in the main setting, particularly those that require sweeping or pouring many small objects. Together, these results establish coding agents as an important baseline for future work on generalized TAMP.

II. Problem Setting

A. MDPs and Simulator Access

We study finite-horizon, goal-directed Markov decision processes (MDPs). An MDP is a tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, \rho, H \rangle$, where $\mathcal{S}$ and $\mathcal{A}$ are the state and action spaces, P is the transition model, R is a sparse reward function indicating goal achievement, ρ is the initial-state distribution, and H is the horizon. States are fully observed and object-centric: a state maps each typed object to a real-valued feature vector describing properties such as pose, geometry, joint configurations, and velocity [20].

Each environment is represented by an MDP and a task description d. The initial-state distribution ρ generates problem instances with varying object counts, configurations, and geometry. For example, in an object-retrieval task, instances may differ in the number and placement of obstacles surrounding the target object. The state and action spaces, transition model, and reward function are shared.

We assume simulator access through `reset` and `step`: a method can sample an initial state $s_0 \sim \rho$ and execute an action a_t from the current state s_t to obtain $s_{t+1} \sim P(\cdot \mid s_t, a_t)$, together with the reward and termination signal.

B. TAMP Environments

We study TAMP environments that are simplified relative to real-world manipulation, following common practice in TAMP research [2]–[7], [19], [20]. With fully observed, object-centric states and no need for perception or language understanding, the remaining challenge is physical reasoning [20]. Horizons are long, rewards are sparse, and the instance distributions are broad enough that no fixed action sequence works. Most importantly, discrete choices are tightly coupled to continuous geometric, kinematic, and dynamic constraints: which object to manipulate, which tool to use, or which subgoal to pursue determines which motions remain feasible, collision constraints grow with the number of objects, and feasible actions can occupy small regions of the action space.

C. Synthesis and Evaluation

At learning time, given the task description d and simulator access, a synthesis method produces a program for $\mathcal{M}$ within a fixed budget. At step t, the program receives $s_t \in \mathcal{S}$ and returns $a_t \in \mathcal{A}$. Since programs can maintain internal state between steps, we denote the induced policy by $a_t = \pi(h_t)$, where $h_t = (s_0, a_0, \ldots, s_t)$ is the interaction history.

At evaluation time, the program is frozen and the synthesis method is no longer invoked. In particular, any coding agent or LLM used during learning is unavailable. The program is evaluated on initial states $s_0 \sim \rho$ that were intentionally hidden during learning.

Our primary metric is success rate: the probability that the program reaches the goal within the horizon H and a wall-clock limit of τ seconds. We additionally measure per-instance computation time spent during planning and deciding on actions, excluding synthesis and time spent advancing the evaluation environment.

Fig. 2: **Unexpected successful strategies found by the coding agents.** Each row shows one execution from left to right; the annotation describes the unexpected behavior. The strategies come from both the main setting and the + source setting.

III. Generalized TAMP with Coding Agents

A. Coding Agents

We instantiate the synthesis method described in Section II using off-the-shelf coding agents from two providers: *Claude Code* running Opus 5 (high), and *Codex* running GPT-5.6 Sol (medium) and GPT-6 Astra (high). We choose Opus 5 as the model for all baselines, so that methods are compared with the same model. Coding agents integrate frontier LLMs with harnesses that enable them to read files, write programs, and execute arbitrary commands. We therefore run the agents inside a sandboxed Docker container with a separate filesystem and no network access. We test this isolation through red-teaming, including attempts to read environment source code, import forbidden libraries, or reach the host or network.

For our main setting, we keep the sandbox bare: the agents have only a Python interpreter with NumPy and SciPy, forcing them to generate end-to-end programs instead of relying on existing libraries. Each agent receives an initial prompt containing the task description d, which includes the environment name and descriptions of the task, observation and action spaces, and goal. We further explain that the environment can vary in object count, and that solutions must support any valid number of objects. We finally explain the evaluation setting, providing the agent with the time τ it will have to run its programs. We provide no hand-written TAMP predicates, operators, samplers, or skills.

To enable interactive learning, we provide the agents with simulator access to the environment: each receives a class that implements `reset` (ρ), `step` (P), and other helpers, including ones to render states as images. The agents see only a client, while the server with the actual implementation of these functions runs outside the container and is inaccessible to the agents. Using this interface, each agent chooses what to run: it can write and execute custom tests, inspect states, and revise its program based on the results.

We also evaluate an additional + *source* setting, with *Claude Code* running Opus 5 and with *Codex* running GPT-6 Astra, where the environment source code is available inside the container. The agents can inspect the implementation and import helper functions, *e.g.*, inverse kinematics solvers. During synthesis, source access also lets the agent set arbitrary states and otherwise manipulate the simulator directly, providing generative access to the transition model [22]. This setting serves as a reference for how the agents perform with complete knowledge of the environment, so the agents can concentrate on developing behavior with less need to infer how the environment works through interaction.

Each agent implements the programmatic policy π as a class with a `reset` method for episode initialization and a `get_action` method for computing $a_t = \pi(h_t)$. Beyond this, we do not constrain the program to any particular abstractions or solution strategy, such as symbolic representations, planning algorithms, or skills. We also ask the agents to commit the program to a git repository before each test, which records its revisions so that we can later replay each one and trace how the program evolved during synthesis.

B. Baselines

We compare AgenticGenPlan against *TAMP planners* that combine symbolic search with sampling to satisfy continuous constraints [7]; the benchmarks provide planners for 16 of the 28 environments. We follow official implementations [6], [20] for the hand-written predicates, operators, samplers, and motion planners (skills) for generating feasible trajectories. They compute a new plan per evaluation instance.

Furthermore, we re-implement *LLMGenPlan* [21], which we run using Opus 5 with chain-of-thought prompting and thinking disabled, as a representative non-agentic LLM generalized planning method. In this setting, the LLM cannot use tools or access the filesystem. It receives a prompt adapted from the original work to our environment and program interfaces, together with the full source code of the environment, and must write a program implementing the same specifications as in our main setting. As the LLM cannot run code, a fixed pipeline evaluates each program generated by the LLM and returns specific pre-defined feedback (an exception with its traceback, an invalid action, or an unsolved instance with its seed). LLMGenPlan receives the same synthesis budget as the coding agents, but the LLM never chooses what to run, cannot write custom tests, and sees only the pre-defined feedback. We finally report the performance of the first program generated by LLMGenPlan as a *One-shot* baseline, to evaluate the LLM without any kind of refinement loop.

IV. Experiments

We design experiments to answer the following questions about the efficacy and efficiency of AgenticGenPlan:

- **Q1.** Can agents write generalized programs for TAMP?
- **Q2.** What strategies do agents discover?
- **Q3.** What advantage does being "agentic" give?
- **Q4.** How efficient are the synthesized programs?
- **Q5.** Does access to environment source code help?

A. Environment Setup

We select the KinDER benchmark [20] as our primary benchmark. This covers 25 environments, grouped into four families: *Kinematic2D*, *Dynamic2D*, *Kinematic3D*, and *Dynamic3D*. The kinematic environments are free of dynamics, while in the dynamic environments, outcomes depend on contact, velocity, and friction, so successful programs may need behaviors such as sweeping, pouring, and tossing. Nineteen of these environments feature variants with different object counts, including counts beyond those evaluated in the original benchmark. We further include three PDDLStream domains, originally introduced by Garrett et al. [6], where LLMs have been shown ineffective [19]: Packing, Blocked, and Rovers.

For each method, we perform five independent runs per environment. Each coding-agent and LLMGenPlan synthesis run has a budget of $20 in model usage. We evaluate the resulting programs and planners on the same 100 instances per environment, sampled from the same initial-state distribution ρ used during synthesis. We generate the evaluation seeds randomly to make it unlikely that agents test them during synthesis, and verify afterward that none were used. We use a timeout τ of 60 seconds per evaluation instance for all methods, matching the original KinDER protocol.

B. Results and Analysis

Tables I and II report success rates across the 28 environments. Below, we write *Opus*, *Sol*, and *Astra* for AgenticGenPlan with *Claude Code* running Opus 5 and *Codex* running GPT-5.6 Sol and GPT-6 Astra, respectively, and *Opus* + source and *Astra* + source for AgenticGenPlan + source. *Astra* is the best performing agent, averaging 99% success on Kinematic2D, 97% on Dynamic2D, 93% on Kinematic3D, 65% on the harder Dynamic3D, and nearly

Method	Kinematic2D						Dynamic2D				Kinematic3D				
	StickButton	Obstruction	ClutteredStorage	ClutteredRetrieval	Motion	PushPullHook	Obstruction	PushPullHook	PushT	ScoopPour	Obstruction	Packing	Transport	Table	BaseMotion
Planner	0.36 [0.36–0.36]	0.41 [0.41–0.41]	0.19 [0.19–0.19]	0.51 [0.51–0.51]	0.71 [0.68–0.73]	–	0.24 [0.24–0.24]	0.13 [0.13–0.13]	–	–	–	0.67 [0.66–0.67]	0.63 [0.59–0.69]	–	**1.00** [1.00–**1.00**]
One-shot Opus 5	0.11 [0.03–0.16]	0.20 [0.00–0.54]	0.02 [0.00–0.06]	0.13 [0.02–0.23]	0.81 [0.17–**1.00**]	0.01 [0.00–0.04]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.17 [0.01–0.36]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.06 [0.00–0.30]	0.00 [0.00–0.00]
LLMGenPlan Opus 5	0.83 [0.70–0.97]	0.82 [0.53–**1.00**]	0.02 [0.00–0.06]	0.32 [0.21–0.41]	0.97 [0.95–**1.00**]	0.12 [0.02–0.34]	0.49 [0.36–0.77]	0.00 [0.00–0.00]	0.17 [0.00–0.41]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.14 [0.00–0.58]	0.05 [0.00–0.27]	0.85 [0.41–**1.00**]	**1.00** [1.00–**1.00**]
AgenticGenPlan GPT-5.6 Sol (medium)	0.94 [0.90–0.99]	0.96 [0.93–0.99]	0.37 [0.19–0.65]	0.21 [0.12–0.27]	**1.00** [1.00–**1.00**]	0.69 [0.46–0.90]	0.77 [0.60–0.90]	0.52 [0.37–0.64]	0.76 [0.21–**1.00**]	0.83 [0.20–**1.00**]	0.26 [0.00–0.69]	0.37 [0.00–0.65]	0.17 [0.00–0.53]	0.99 [0.97–**1.00**]	**1.00** [1.00–**1.00**]
AgenticGenPlan Opus 5 (high)	**1.00** [1.00–**1.00**]	**1.00** [1.00–**1.00**]	0.99 [0.99–**1.00**]	0.81 [0.36–0.98]	**1.00** [1.00–**1.00**]	0.98 [0.91–**1.00**]	0.88 [0.79–0.94]	0.83 [0.63–0.95]	**1.00** [1.00–**1.00**]	0.97 [0.89–**1.00**]	0.85 [0.41–**1.00**]	0.66 [0.52–0.79]	**0.99** [0.94–**1.00**]	**1.00** [1.00–**1.00**]	**1.00** [1.00–**1.00**]
AgenticGenPlan GPT-6 Astra (high)	**1.00** [1.00–**1.00**]	1.00 [0.99–**1.00**]	**1.00** [0.99–**1.00**]	**0.96** [0.91–**1.00**]	**1.00** [1.00–**1.00**]	**1.00** [1.00–**1.00**]	**0.96** [0.94–**0.98**]	**0.91** [0.79–**1.00**]	1.00 [0.99–**1.00**]	**1.00** [0.99–**1.00**]	**0.97** [0.92–0.99]	**0.89** [0.67–**1.00**]	0.78 [0.00–**1.00**]	**1.00** [1.00–**1.00**]	**1.00** [1.00–**1.00**]
AgenticGenPlan + source Opus 5 (high)	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	0.96 [0.92–0.99]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	0.92 [0.87–0.96]	0.92 [0.77–0.99]	1.00 [1.00–1.00]	0.88 [0.65–1.00]	1.00 [0.99–1.00]	0.98 [0.94–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]
AgenticGenPlan + source GPT-6 Astra (high)	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	0.93 [0.83–0.98]	0.97 [0.91–1.00]	1.00 [1.00–1.00]	0.97 [0.87–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]

TABLE I: **Success rate over environments**. We report the mean over five runs, with [min–max] across run-level success rates below. Bold marks the best mean and the best max per environment among the main-setting rows. AgenticGenPlan + source additionally gives the agent the environment source code. A dash marks environments for which KinDER provides no planner. Dynamic3D and PDDLStream environments are listed in Table II.

Method	Dynamic3D										PDDLStream		
	BalanceBeam	ConstrainedCupboard	Dynamo	Rearrange	ScoopPour	Shelf	SortClutteredBlocks	SweepIntoDrawer	SweepSimple	Tossing	Packing	Blocked	Rovers
Planner	–	–	–	–	–	0.27 [0.22–0.30]	–	0.00 [0.00–0.00]	–	0.77 [0.77–0.78]	0.54 [0.50–0.56]	0.74 [0.70–0.77]	0.30 [0.24–0.38]
One-shot Opus 5	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.13 [0.01–0.34]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.12 [0.00–0.40]	0.00 [0.00–0.01]	0.26 [0.00–0.87]
LLMGenPlan Opus 5	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.98 [0.95–**1.00**]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.01]	0.21 [0.00–0.39]	0.00 [0.00–0.01]	0.86 [0.81–0.97]
AgenticGenPlan GPT-5.6 Sol (medium)	0.64 [0.00–**1.00**]	0.00 [0.00–0.00]	**1.00** [1.00–**1.00**]	0.09 [0.00–0.32]	0.00 [0.00–0.00]	0.24 [0.00–0.63]	0.01 [0.00–0.03]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	0.81 [0.41–**1.00**]	0.75 [0.34–0.99]	0.78 [0.00–**1.00**]
AgenticGenPlan Opus 5 (high)	0.97 [0.87–**1.00**]	0.08 [0.00–0.34]	0.92 [0.60–**1.00**]	0.58 [0.01–0.96]	0.09 [0.00–**0.31**]	0.60 [0.00–**1.00**]	0.30 [0.13–**0.72**]	0.00 [0.00–0.00]	0.00 [0.00–0.00]	**0.99** [0.96–**1.00**]	0.99 [0.96–**1.00**]	0.38 [0.25–0.56]	0.98 [0.90–**1.00**]
AgenticGenPlan GPT-6 Astra (high)	**1.00** [0.99–**1.00**]	**0.29** [0.04–**0.63**]	**1.00** [1.00–**1.00**]	**0.96** [0.93–**0.98**]	**0.14** [0.00–0.23]	**1.00** [0.98–**1.00**]	**0.43** [0.31–0.50]	**0.72** [0.15–**1.00**]	**0.03** [0.00–**0.15**]	0.96 [0.82–**1.00**]	**1.00** [1.00–**1.00**]	**0.99** [0.98–**1.00**]	**1.00** [0.99–**1.00**]
AgenticGenPlan + source Opus 5 (high)	1.00 [0.99–1.00]	0.47 [0.05–1.00]	1.00 [1.00–1.00]	0.77 [0.15–0.97]	0.38 [0.21–0.57]	0.43 [0.00–0.98]	0.49 [0.20–0.96]	0.57 [0.00–0.97]	0.07 [0.00–0.21]	0.85 [0.31–1.00]	1.00 [1.00–1.00]	0.87 [0.43–1.00]	0.99 [0.98–1.00]
AgenticGenPlan + source GPT-6 Astra (high)	1.00 [0.99–1.00]	1.00 [0.99–1.00]	1.00 [1.00–1.00]	0.99 [0.97–1.00]	0.42 [0.06–1.00]	1.00 [0.99–1.00]	0.95 [0.89–0.99]	0.86 [0.47–1.00]	0.44 [0.35–0.50]	0.98 [0.93–1.00]	1.00 [0.98–1.00]	1.00 [1.00–1.00]	1.00 [1.00–1.00]

TABLE II: **Dynamic3D and PDDLStream environments**. Same protocol and notation as Table I. The planner for the PDDLStream environments is PDDLStream itself.

100% on PDDLStream. *Opus* follows, with 96%, 92%, 90%, 45%, and 78%, respectively. All three agent configurations exceed the planning baseline in most environments with one available (15 of 16 for *Astra*, 12 for *Opus*, and 9 for *Sol*), without the hand-designed models, skills, and samplers that the planners use, and their programs exploit regularities within each environment to restrict the decisions considered at test time (**Q1**).

Discovering Unexpected Strategies: We find that the coding agents discover unexpected manipulation strategies, depicted in Figure 2 (**Q2**). To name a few, in Dynamic2D ScoopPour (row three), an *Opus* program in the main setting rotates the tool and regrasps it from the left side, which lets it scoop far more balls at once. In SweepIntoDrawer (row five), an *Opus* + source program ignores the tool (sweeper) and uses the gripper to sweep the cubes one by one.

Testing Edge Cases: One advantage that agentic methods provide over static LLM calls is the flexibility to write and execute custom tests (**Q3**). These let agents investigate failures and evaluate their policies on selected configurations. In StickButton, for example, *Opus* writes a script that repeatedly calls `reset` with different seeds, inspects the sampled states, and saves edge cases involving wall-adjacent sticks and high buttons. It combines these edge cases with typical instances to build a diverse test suite, then tests its programs on both to challenge them across a broad range of configurations (Figure 3, left).

Building Internal Models: Compared to static LLM-based synthesis methods, we also find that AgenticGenPlan draws on prior robotics knowledge to build initial models, then tests and calibrates them through interaction (**Q3**). In Shelf, for example, one *Opus* run constructs an initial kinematic model

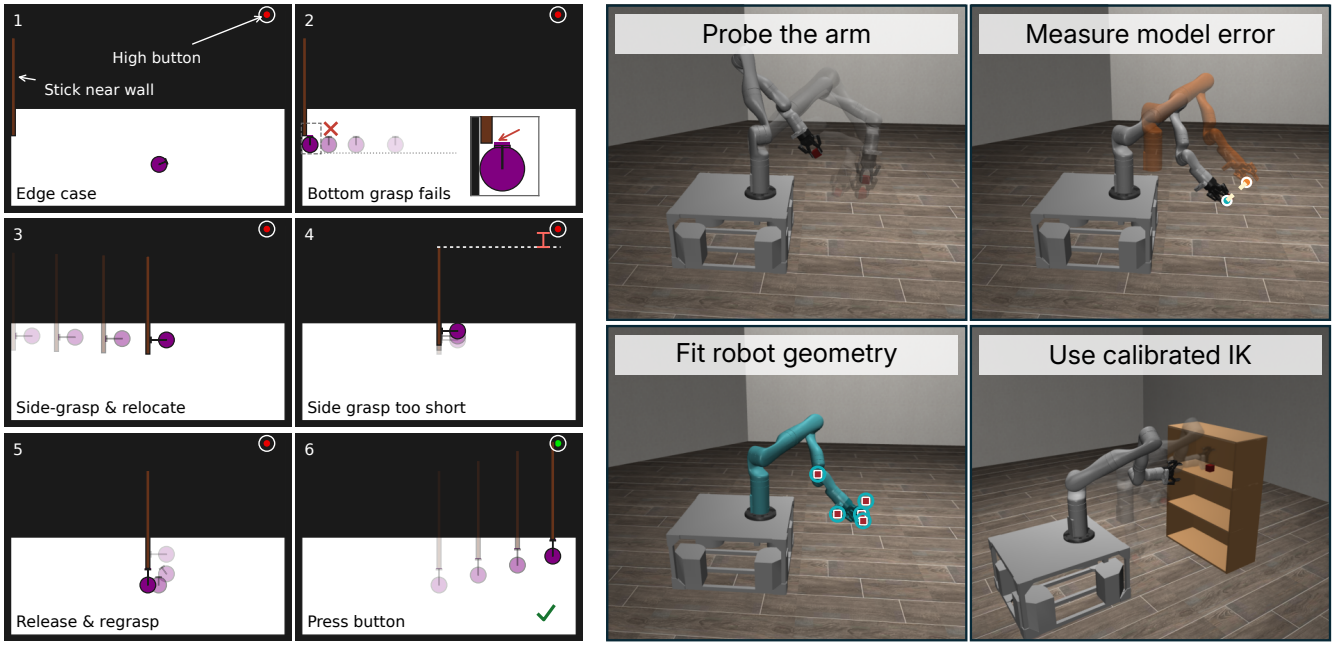

Fig. 3: **Left: custom tests reveal edge cases in StickButton.** *Opus* writes a script that repeatedly calls `reset` with different seeds to find wall-adjacent sticks and high buttons for testing its program. The wall prevents a bottom grasp. A side grasp lets the robot move the stick away from the wall but leaves the high button out of reach; releasing and re-grasping from below provides the required height. **Right: learning robot geometry through interaction.** In Shelf, *Opus* probes the arm while holding a cube, fits its kinematic model to observed cube positions, and uses the calibrated model for inverse kinematics (IK). The initial model discrepancy is exaggerated for visibility.

of the Kinova Gen3 arm from the agent's own knowledge, without internet access. It then uses a grasped cube as a marker because the state exposes object positions but not the hand's position. Moving the arm through different configurations lets it calibrate the model's predictions of cube positions from joint angles (Figure 3, right). From a few observations, it fits six parameters describing the robot mount and grasp offsets, reducing the RMSE between predicted and observed cube positions from 38.9 to 1.8 mm on the calibration observations. It later uses the fitted model to solve inverse kinematics (IK) as part of the solution.

Building on what they learn about the environment, agents continue interacting with it to refine control parameters and manipulation strategies. In BalanceBeam, for example, one *Sol* run moves its small-block placement targets closer to the beam's center, increasing success from 6% to 64%.

Comparing Agents: *Astra* achieves the highest mean success, above *Opus* in 20 of 28 environments. In 11 of these 20, the best *Opus* program scores at least as high as the best *Astra* program, so the gain comes from greater consistency: in Shelf, for example, *Opus* programs range from 0% to 100% success. The weakest never discovers which shelf the cubes must go on, and another places one cube on the correct shelf but leaves the rest on the floor in front of the shelf. In contrast, every *Astra* program is tested and revised on instances with up to eight cubes and scores at least 98%. In other environments, *Astra* finds strategies that no *Opus* program uses. In SweepIntoDrawer, where no *Opus* program succeeds, the best *Astra* program solves every instance by picking up the cubes instead of sweeping them, while the only *Astra* program that sweeps scores 15%. In Kinematic3D Packing, instances with three parts combine a cube with two triangles, each either right or equilateral, on a small rack. The best *Opus* program places right triangles in fixed, hand-tuned positions but has no way to place equilateral ones. Three *Astra* programs instead search over positions and rotations for every part and solve all instances. Some *Opus* programs are still better: in Dynamic3D ScoopPour, all agents skip scooping and move the whole tray at once, and the best *Opus* program tips the tray over while resting it on the target tray, whereas the best *Astra* program flips the tray entirely in mid-air and cannot recover when it drops it.

Between the other two agents, *Opus* is stronger than *Sol*, achieving higher mean success in 22 of 28 environments, with four ties. *Sol* is higher only in Dynamo and in Blocked, where it reaches 75%, compared with 38% for *Opus* and 74% for the planner. In Blocked, the robot can retrieve either a nearby obstructed block or, when available, an unobstructed block on a distant table. *Opus* programs often persist with the nearby block, where obstacle orientations make grasping and retrieval difficult. *Sol* programs can instead switch to the distant block after unsuccessful attempts. In one *Sol* run, a commit adding this fallback raises success from 15% to 56%, with 41 newly solved instances and no lost successes (Figure 4). Subsequent commits reach 83% final success, solving all 80 instances with an alternative block but only three of the 20 without one. *Astra* programs combine both behaviors: one run grasps the nearby block reliably, even at awkward orientations, and falls back to the distant table when a grasp fails, solving all 100 instances.

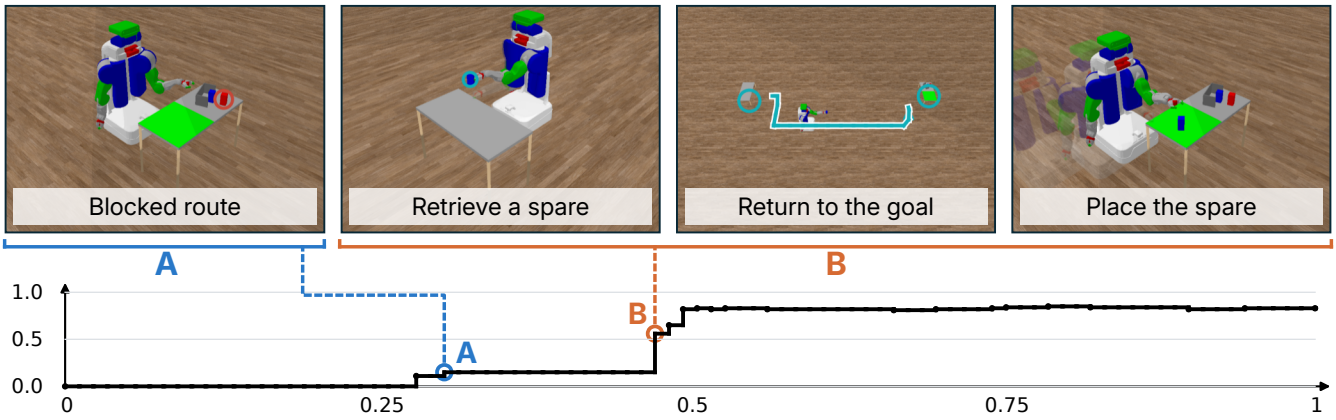

Fig. 4: **Refining manipulation strategies during synthesis.** In one *Sol* Blocked run, a revision adding a spare-block fallback changes success from 15% (A) to 56% (B); the pictured execution retrieves an alternative goal block from a distant table. The curve shows the held-out success rate of each commit, in commit order.

Method	Time/action (ms)
AgenticGenPlan Opus 5 (high)	11.741 [0.041–109.403]
AgenticGenPlan + source Opus 5 (high)	42.480 [0.056–381.509]
AgenticGenPlan GPT-6 Astra (high)	1.299 [0.033–6.047]
AgenticGenPlan + source GPT-6 Astra (high)	50.392 [0.053–477.045]

TABLE III: **Policy-computation time per action** on the 15 environments where *Opus*, *Opus* + source, *Astra*, and *Astra* + source each reach 100% held-out success in at least one run. Each method is averaged over its own perfect runs within each environment; we report the mean across environments, with [min–max] across environment-level means. Source access yields slower programs on average, as many of them invoke the environment source code as part of planning at decision time.

Program Computation Efficiency: Table III compares policy-computation time per action on the 15 environments where *Opus*, *Opus* + source, *Astra*, and *Astra* + source each reach 100% held-out success in at least one run, averaging each method over its own perfect runs (**Q4**). Main-setting *Opus* programs choose actions quickly, averaging 11.7 ms per action. *Opus* + source programs average over 3 times that, with large variation across environments, from over twice as fast to nearly 200 times slower, because many of them invoke the environment source code as part of planning. Synthesis without source code instead forces the agent to distill what it learns into self-contained code, which yields faster programs at evaluation time. *Astra* programs are the fastest, averaging 1.3 ms per action, about 9 times faster than *Opus*, and source access costs them even more, raising the average to 50.4 ms. *Astra* programs tend to use compact, closed-form control, such as a few hardcoded joint poses or analytic inverse kinematics, whereas *Opus* programs more often search over large precomputed candidate sets, with retry and fallback logic, at decision time. Programs also need far less computation than the planners: on the 14 environments with a planner available and multiple object counts, *Opus* and *Astra* programs take 2.1 s and 0.5 s per instance on average, compared with 29 s for the planners.

Environment Source Code Access: We further inspect whether providing the agents with environment source code helps them generate programs (**Q5**). Source access raises mean success for both agents, from 74% to 84% for *Opus* and from 86% to 95% for *Astra*. Source access is not sufficient on its own: LLMGenPlan also receives the source code, uses Opus 5, and has the same $20 budget, yet reaches 28%, while *Opus* + source is higher in 27 of 28 environments. Most of this gap points to the benefit of agentic synthesis, where the agent can run code, test, and revise (**Q3**). Environment source code provides both information and implementations that programs can reuse. With source access, the agent can read goals and success checks from the code, whereas in the main setting the agent must infer them from rewards and rendered images, and sometimes settles on an inaccurate goal and writes its program against it. In SortClutteredBlocks, for example, main-setting *Astra* programs assign cubes to bins by their order in the state, with a rule that works for four cubes but fails on every twenty-cube instance. With source access, programs read each cube's target bin (43% to 95%). Programs also reuse the implementation directly, for example loading the robot's model for inverse kinematics, calling motion planners and collision checks, or testing actions in private simulators. In Kinematic3D Packing, one *Opus* run uses internal collision information to identify that the gripper, rather than the held part, collides with the rack. It then changes the grasp to provide clearance. In SweepIntoDrawer, where *Opus* and *Sol* score zero in the main setting, source access raises *Opus*'s mean success to 57%, with one run reaching 97%.

V. Related Work

A. Generalized Planning and Learning for TAMP

TAMP couples discrete decisions with continuous feasibility constraints [7]. Systems such as PDDLStream provide interfaces between symbolic planning and procedures for sampling and checking continuous quantities [6]. Generalized planning seeks solutions that apply across related problem instances [8]. In TAMP, this objective has motivated learning

reusable samplers, feasibility predictors, search heuristics, and state and action abstractions [9], [11], [13], [14]; see [10] for a recent survey. We ask whether general-purpose coding agents can likewise exploit regularities across instances without being confined by specific TAMP abstractions.

B. Foundation Models for Task and Motion Planning

One branch of foundation model-driven TAMP research leverages large language models (LLMs) to guide a TAMP planner. For example, Text2Motion combines language-guided task planning with skill affordances [23], and LLM^3 uses motion-planning failures to guide plan revision [24]. In addition to guidance, LLM-based program synthesis can also automate the engineering of TAMP components. PRoC3S generates programs whose continuous parameters are resolved through constraint satisfaction [25]. MOPS searches over programs specifying constraints for trajectory optimization [26], while OWL-TAMP uses vision-language models to generate constraints within a TAMP system [27]. Instead of committing to certain TAMP abstractions, our evaluation leaves the internal structure of the generated solution open: an agent may implement search, optimization, sampling, or task-specific procedures on its own, with the final objective of efficiently solving as many problem instances as possible.

C. Agentic Synthesis of Robotic Programs

Since Code-as-Policies [28], LLMs have demonstrated the ability to generate robot control code, integrating manipulation skills and perception interfaces. Later works such as GenCHiP [29] and InstructFlow [30] motivate code as a flexible policy representation, while highlighting the importance of the supplied action interface and feedback. In symbolic task planning, Silver et al. [21] use LLMs and execution feedback to synthesize reusable programs that solve new PDDL instances without further LLM calls.

Recent agentic systems provide even closer precedents for reusable policy synthesis. RHO introduces a reflective evolutionary optimizer over policy repositories, using coding agents and execution feedback [31]. ASPIRE develops mechanisms for discovering, repairing, and reusing robotic skills [32], and MEMENTO introduces memory-guided evolutionary search over policy programs [33]. The benchmarks used in these works, such as Robosuite [34] and LIBERO [35], focus on general-purpose manipulation rather than the physical reasoning challenges of TAMP problems [20]. Our focus is a systematic assessment of off-the-shelf coding agents generating code as generalized policies for constrained TAMP-like environments.

D. Systematic Evaluation of Embodied Agents

Mendez-Mendez's study of LLMs within PDDLStream is the closest precedent in evaluation focus [19]. It examines configurations in which LLMs replace task planning, continuous sampling, or both within established planning architectures, and compares them with engineered planners. Whereas it queries LLMs during problem solving, we use simulator feedback to develop a program, then freeze it for LLM-free evaluation on unseen instances. Several benchmarks and systematic studies also inform our evaluation. CaP-X benchmarks coding agents for general manipulation control [36]. Tsui et al. [37] evaluate an LLM agent SDK on general manipulation tasks with iterative execution. KinDER [20] provides procedurally generated environments designed to isolate physical reasoning and already compares planning and learning approaches; we build on these environments and their baseline implementations.

VI. Discussion

Limitations: Our setup assumes fully observed, object-centric states and simulator access during synthesis. It therefore does not establish performance under perception uncertainty. The models' training data are undisclosed, so prior exposure to benchmark code cannot be excluded. However, the synthesis logs show policies being developed through environment probing, testing, and substantial revision. Together with the unexpected strategies the agents synthesized, this provides evidence against simply recalling complete solutions seen during pretraining.

Conclusions: We show that coding agents are strong generalized TAMP planners, synthesizing reusable programs that outperform hand-engineered planners and existing LLM-based synthesis methods on our benchmarks. These results extend the promise of LLM-based generalized planning to environments with geometric, kinematic, and dynamic constraints. Through interaction, agents can investigate the physical behavior of an environment and develop effective strategies without being supplied with its symbolic models, skills, or samplers. This ability to discover both how an environment works and how to solve its tasks makes coding agents an important baseline for future TAMP research.

Acknowledgments

We thank Pietro Ferrazzi for feedback on a draft of this paper. This work was partially supported by a Princeton SEAS Innovation Grant, an NVIDIA Academic Grant Program, and a Princeton AI Lab Grant.

References

[1] A. Deshpande, "Exact geometry algorithms for robotic motion planning," Ph.D. dissertation, Massachusetts Institute of Technology, 2019. [Online]. Available: https://hdl.handle.net/1721.1/122736

[2] L. P. Kaelbling and T. Lozano-Pérez, "Hierarchical task and motion planning in the now," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011. [Online]. Available: https://doi.org/10.1109/icra.2011.5980391

[3] S. Srivastava, E. Fang, L. Riano, R. Chitnis, S. Russell, and P. Abbeel, "Combined task and motion planning through an extensible planner-independent interface layer," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014. [Online]. Available: https://doi.org/10.1109/icra.2014.6906922

[4] M. Toussaint, "Logic-geometric programming: An optimization-based approach to combined task and motion planning," in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015. [Online]. Available: https://www.ijcai.org/Abstract/15/274

[5] N. T. Dantam, Z. K. Kingston, S. Chaudhuri, and L. E. Kavraki, "An incremental constraint-based framework for task and motion planning," *The International Journal of Robotics Research*, vol. 37, no. 10, pp. 1134–1151, 2018. [Online]. Available: https://doi.org/10.1177/0278364918761570

[6] C. R. Garrett, T. Lozano-Pérez, and L. P. Kaelbling, "PDDLStream: Integrating symbolic planners and blackbox samplers via optimistic adaptive planning," in *International Conference on Automated Planning and Scheduling (ICAPS)*, 2020. [Online]. Available: https://doi.org/10.1609/icaps.v30i1.6739

[7] C. R. Garrett, R. Chitnis, R. Holladay, B. Kim, T. Silver, L. P. Kaelbling, and T. Lozano-Pérez, "Integrated task and motion planning," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 265–293, 2021. [Online]. Available: https://doi.org/10.1146/annurev-control-091420-084139

[8] S. Jiménez, J. Segovia-Aguas, and A. Jonsson, "A review of generalized planning," *The Knowledge Engineering Review*, vol. 34, 2019. [Online]. Available: https://doi.org/10.1017/s0269888918000231

[9] A. Curtis, T. Silver, J. B. Tenenbaum, T. Lozano-Pérez, and L. P. Kaelbling, "Discovering state and action abstractions for generalized task and motion planning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022. [Online]. Available: https://doi.org/10.1609/aaai.v36i5.20475

[10] Y. Huang, B. Hedegaard, N. Shah, S. Srivastava, G. Konidaris, and T. Silver, "Learning by and for task and motion planning: A survey," 2026, manuscript. https://prpl-group.com/tamp-learning-survey.pdf.

[11] B. Kim, L. P. Kaelbling, and T. Lozano-Pérez, "Guiding search in continuous state-action spaces by learning an action sampler from off-target search experience," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. [Online]. Available: https://doi.org/10.1609/aaai.v32i1.12106

[12] Z. Wang, C. R. Garrett, L. P. Kaelbling, and T. Lozano-Pérez, "Learning compositional models of robot skills for task and motion planning," *The International Journal of Robotics Research*, vol. 40, no. 6–7, pp. 866–894, 2021. [Online]. Available: https://doi.org/10.1177/02783649211004615

[13] A. M. Wells, N. T. Dantam, A. Shrivastava, and L. E. Kavraki, "Learning feasibility for task and motion planning in tabletop environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1255–1262, 2019. [Online]. Available: https://doi.org/10.1109/lra.2019.2894861

[14] R. Chitnis, D. Hadfield-Menell, A. Gupta, S. Srivastava, E. Groshev, C. Lin, and P. Abbeel, "Guided search for task and motion plans using learned heuristics," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016. [Online]. Available: https://doi.org/10.1109/icra.2016.7487165

[15] C. E. Jimenez, J. Yang, A. Wettig, S. Yao, K. Pei, O. Press, and K. Narasimhan, "SWE-bench: Can language models resolve real-world GitHub issues?" in *International Conference on Learning Representations (ICLR)*, 2024. [Online]. Available: https://openreview.net/forum?id=VTF8yNQM66

[16] J. Yang, C. E. Jimenez, A. Wettig, K. Lieret, S. Yao, K. Narasimhan, and O. Press, "SWE-agent: Agent-computer interfaces enable automated software engineering," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2024/hash/5a7c947568c1b1328ccc5230172e1e7c-Abstract-Conference.html

[17] A. B. Corrêa, A. G. Pereira, and J. Seipp, "Classical planning with LLM-generated heuristics: Challenging the state of the art with Python code," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2025/hash/3bf4b55960aaa23553cd2a6bdc6e1b57-Abstract-Conference.html

[18] A. B. Corrêa, A. G. Pereira, and J. Seipp, "Frontier large language models rival state-of-the-art planners," *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2511.09378

[19] J. Mendez-Mendez, "A systematic study of large language models for task and motion planning with PDDLStream," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2026. [Online]. Available: https://arxiv.org/abs/2510.00182

[20] Y. Huang, B. Li, V. Saxena, Y. Liang, U. A. Mishra, L. Ji, L. Zha, J. Wu, N. Kumar, S. Scherer, D. Xu, and T. Silver, "KinDER: A physical reasoning benchmark for robot learning and planning," in *Robotics: Science and Systems (RSS)*, 2026. [Online]. Available: https://doi.org/10.15607/rss.2026.xxii.184

[21] T. Silver, S. Dan, K. Srinivas, J. B. Tenenbaum, L. P. Kaelbling, and M. Katz, "Generalized planning in PDDL domains with pretrained large language models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024. [Online]. Available: https://doi.org/10.1609/aaai.v38i18.30006

[22] M. Kearns, Y. Mansour, and A. Y. Ng, "A sparse sampling algorithm for near-optimal planning in large Markov decision processes," *Machine Learning*, vol. 49, no. 2–3, pp. 193–208, 2002. [Online]. Available: https://doi.org/10.1023/A:1017932429737

[23] K. Lin, C. Agia, T. Migimatsu, M. Pavone, and J. Bohg, "Text2Motion: From natural language instructions to feasible plans," *Autonomous Robots*, vol. 47, pp. 1345–1365, 2023. [Online]. Available: https://doi.org/10.1007/s10514-023-10131-7

[24] S. Wang, M. Han, Z. Jiao, Z. Zhang, Y. N. Wu, S.-C. Zhu, and H. Liu, "LLM^3: Large language model-based task and motion planning with motion failure reasoning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024. [Online]. Available: https://doi.org/10.1109/iros58592.2024.10801328

[25] A. Curtis, N. Kumar, J. Cao, T. Lozano-Pérez, and L. P. Kaelbling, "Trust the PRoC3S: Solving long-horizon robotics problems with LLMs and constraint satisfaction," in *Proceedings of the Conference on Robot Learning (CoRL)*, vol. 270, 2025, pp. 1362–1383. [Online]. Available: https://proceedings.mlr.press/v270/curtis25a.html

[26] D. Shcherba, E. Cobo-Briesewitz, C. V. Braun, and M. Toussaint, "Meta-optimization and program search using language models for task and motion planning," in *Proceedings of the Conference on Robot Learning (CoRL)*, vol. 305, 2025, pp. 5339–5361. [Online]. Available: https://proceedings.mlr.press/v305/shcherba25a.html

[27] N. Kumar, W. Shen, F. Ramos, D. Fox, T. Lozano-Pérez, L. P. Kaelbling, and C. R. Garrett, "Open-world task and motion planning via vision-language model generated constraints," *IEEE Robotics and Automation Letters*, vol. 11, no. 3, pp. 3366–3373, 2026. [Online]. Available: https://doi.org/10.1109/lra.2026.3656799

[28] J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence, and A. Zeng, "Code as policies: Language model programs for embodied control," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 9493–9500. [Online]. Available: https://doi.org/10.1109/icra48891.2023.10160591

[29] K. Burns, A. Jain, K. Go, F. Xia, M. Stark, S. Schaal, and K. Hausman, "GenCHiP: Generating robot policy code for high-precision and contact-rich manipulation tasks," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 9596–9603. [Online]. Available: https://doi.org/10.1109/iros58592.2024.10801525

[30] H. Chi, Z. Feng, Y. Lyu, C. Zheng, L. Luo, Y. S. Ong, I. Tsang, H. Chen, Y. Chang, and H. Yin, "InstructFlow: Adaptive symbolic constraint-guided code generation for long-horizon planning," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 38, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2025/hash/03cfff3eccb29aa15f76e9bcee3d1be7-Abstract-Conference.html

[31] K. Elmaaroufi, J. Svegliato, S. Kalade, G. Schelle, S. A. Seshia, and M. Zaharia, "RHO: Your coding agent is secretly a roboticist," arXiv preprint, 2026. [Online]. Available: https://arxiv.org/abs/2606.16458

[32] R. Lu, Y. Wu, E. Kou, L. Fu, W. Xiao, A. Mandlekar, Y. Xu, G. Shi, K. Goldberg, A. Chen, M. Chowdhury, Y. Zhu, L. Fan, and G. Wang, "ASPIRE: Agentic /skills discovery for robotics," arXiv preprint, 2026. [Online]. Available: https://arxiv.org/abs/2607.00272

[33] A. Sygkounas, V. Aregbede, A. Loutfi, and A. Persson, "MEMENTO: Memory-guided memetic code-as-policy evolution," arXiv preprint, 2026. [Online]. Available: https://arxiv.org/abs/2607.22832

[34] Y. Zhu, J. Wong, A. Mandlekar, R. Martín-Martín, A. Joshi, K. Lin, A. Maddukuri, S. Nasiriany, and Y. Zhu, "robosuite: A modular simulation framework and benchmark for robot learning," arXiv preprint, 2020. [Online]. Available: https://arxiv.org/abs/2009.12293

[35] B. Liu, Y. Zhu, C. Gao, Y. Feng, Q. Liu, Y. Zhu, and P. Stone, "LIBERO: Benchmarking knowledge transfer for lifelong robot learning," in *Advances in Neural Information Processing Systems*, vol. 36, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/hash/8c3c666820ea055a77726d66fc7d447f-Abstract-Datasets_and_Benchmarks.html

[36] L. Fu, J. Yu, K. El-Refai, E. Kou, H. Xue, H. Huang, W. Xiao, G. Wang, D. Niu, F.-F. Li, G. Shi, J. Wu, S. Sastry, Y. Zhu, K. Goldberg, and L. Fan, "CaP-X: A framework for benchmarking and improving coding agents for robot manipulation," arXiv preprint, 2026. [Online]. Available: https://arxiv.org/abs/2603.22435

[37] B. Y. Tsui, A. Y. Fang, and T. J. Hwu, "Demonstration-free robotic control via LLM agents," arXiv preprint, 2026. [Online]. Available: https://arxiv.org/abs/2601.20334