

Weights or Skills?

A Survey of Robot-Learning Techniques: from Action-Predicting Weights to Robots that Write their Own Skills

GAYTRI JENA*, UC Berkeley, USA

KAPIL WANASKAR, San Jose State University, USA

VINIJA JAIN, Meta, USA

AMAN CHADHA, Apple, USA

VASU SHARMA, PocketFM, USA

AMITAVA DAS, Pragya Lab, BITS Pilani Goa, India

Robot learning is splitting into two bets: policies that bake competence into *frozen weights* (vision-language-action, or VLA, models), and agents that write and refine their own *executable skills* as code. This survey organises the field around that axis of **weights versus skills**. Its central analytical contribution is a deep-dive that arranges code-as-policy methods by their *degree of self-improvement*, from zero-shot program synthesis, through closed-loop self-repair and persistent skill memory, to the sparsely populated cell in which execution feedback, skill memory, and evolutionary search combine into one open-ended loop; only a few very recent systems (for example ASPIRE, ENPIRE, and RoboClaw) occupy that cell. We map the complementary “skills” pole, from unsupervised reinforcement-learning skill discovery to large-language-model skill libraries, and show that the word “skill” is used in at least five distinct senses, of which only the code sense self-improves without gradient updates. We then connect the taxonomy to the emerging *skill economy*: commercial robot-skill marketplaces now distribute one-tap skills across robots but ship only static playback, which surfaces open problems of adaptation, cross-embodiment portability, provenance, safety verification, composition, and standardisation. This is a deliberately *focused* survey. Rather than cataloguing the field exhaustively, it examines 77 representative systems across six technique families through one taxonomy and a set of contrast tables, and it supplies operational definitions of the self-improvement mechanisms together with a statement of what each family cannot do.

CCS Concepts: • **Computing methodologies** → **Robotic planning**; *Machine learning*.

Additional Key Words and Phrases: robot learning, code-as-policy, self-improvement, vision-language-action models, skill libraries, physical AI, survey

1 INTRODUCTION

Two paradigms now dominate robot learning. *Vision-language-action* (VLA) models predict actions from frozen weights; *code-as-policy* agents instead write executable programs and improve them from experience [144, 160]. This survey is organised around that fault line. Figure 1 poses the question with real systems: frozen-weight VLA policies on one side, executable and self-improving code-as-policy agents on the other.

Why now: the skill stack. The app store for robot skills has arrived; Unitree UniStore ships one-tap, cross-model motion downloads [243]. But every skill in it is *static playback*: the least-capable point in the taxonomy. The skill stack has three layers: (1) **Build** (how skills are created, §2–§3), (2) **Distribute** (marketplaces and hubs), and (3) **Adapt** (on-device self-improvement, §3.1). UniStore is layer 2 over static layer-1 skills; the promised value, “adjust grip per fruit so nothing bruises”, lives in layer 3, which no shipped skill does today.

*The authors contributed to this work independently of their roles and employment.

Authors’ addresses: Gaytri Jena, UC Berkeley, Berkeley, CA, USA; Kapil Wanaskar, San Jose State University, USA; Vinija Jain, Meta, USA; Aman Chadha, Apple, USA; Vasu Sharma, PocketFM, USA; Amitava Das, Pragya Lab, BITS Pilani Goa, India.

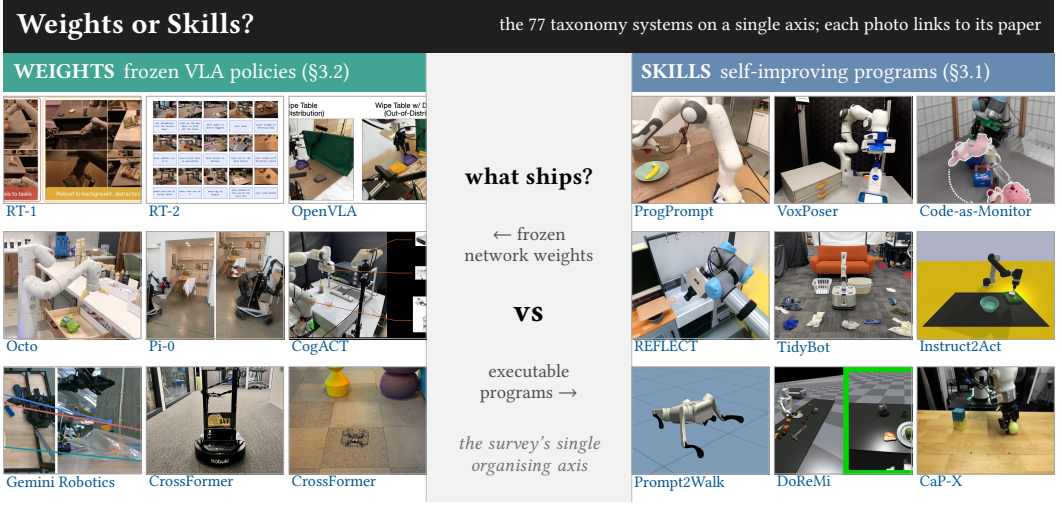


Fig. 1. **Weights or Skills?** The survey’s organising question, posed with real systems. *Weights* (left): vision-language-action policies ship frozen network weights that map observations to actions (§3.2). *Skills* (right): code-as-policy agents ship executable programs and improve them from their own experience (§3.1). The 77 taxonomy systems populate these two poles across six technique branches (2016–2026). Every photo is a hyperlink to its source paper; per-panel provenance is in Table 10. Images © their respective authors, reproduced for scholarly review.

Scope and contributions. This is a *focused* survey rather than an exhaustive census. We examine 77 systems chosen, by the explicit placement criteria of §2.1, to populate a single analytical taxonomy, and we exclude pure perception, navigation, and locomotion work except where it bears directly on how a manipulation skill is authored and improved. Our contributions are: (i) a taxonomy organised by *weights vs. skills* (§2); (ii) a deep-dive that arranges code-as-policy by *degree of self-improvement*, with operational definitions of the feedback, memory, and search mechanisms and necessary-and-sufficient conditions for each rung (§3.1); (iii) a map of the “skills” pole together with an analysis of the five senses in which “skill” is used (§3.4); (iv) a summary of the characteristic strengths and limitations of each technique family; and (v) an open-problems agenda for the emerging skill economy (§4). To situate these systems within the wider field, Appendix A (Table 11) catalogues 225 further representative works, compared on six axes and grouped by area, bringing the survey’s total coverage to more than three hundred works.

Related surveys. Table 1 contrasts the topical coverage of the closest surveys with ours, and Figure 3 places them on a timeline. Within *ACM Computing Surveys* the nearest are on world models, embodied intelligence, and language-model navigation, and beyond it a cluster covers vision-language-action models and foundation models for manipulation; but none organises code-as-policy by degree of self-improvement or names the §3.1.5 cell.

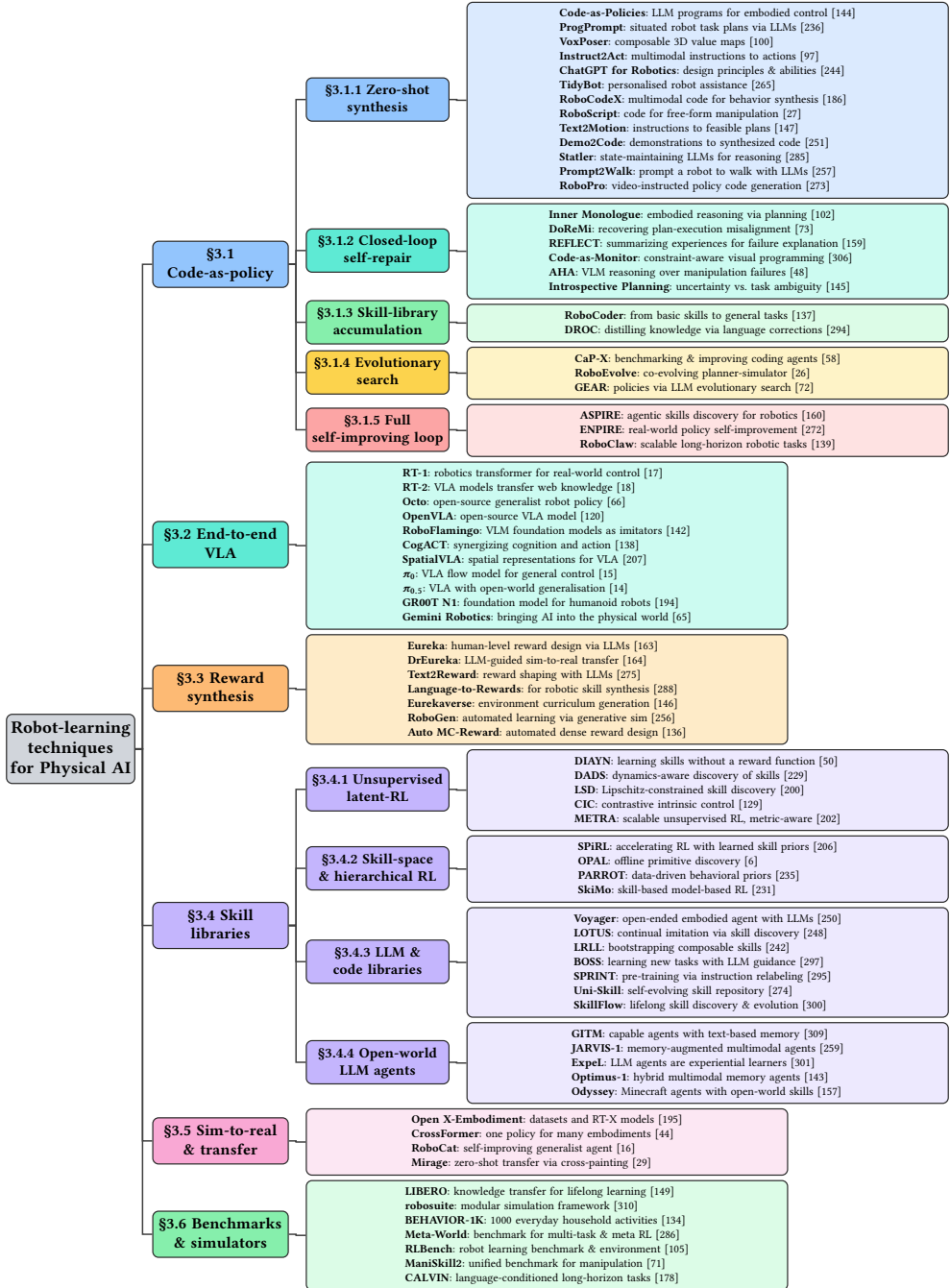


Fig. 2. **The weights-versus-skills taxonomy of robot learning** (all 77 systems). Each sub-family is one cell listing its systems, each with a trimmed form of the paper’s own title. Left fork = inspectable *code* / *skills* (§3.1, §3.4); the §3.1 cells are shaded by *degree of self-improvement*, from zero-shot to the full feedback+memory+search loop. Full per-system detail: Tables 2–6.

Survey	Venue	Yr	weights	code	skill	reward	Self-impr.	Economy
World Models [41]	CSUR	2025	~	✗	✗	✗	✗	✗
Embodied Intelligence [153]	CSUR	2025	~	✗	~	✗	✗	✗
Navigation via FLM [197]	CSUR	2026	~	~	✗	✗	✗	✗
VLA for Embodied AI [161]	arXiv	2024	✓	✗	✗	✗	✗	✗
Large-VLM VLA [228]	arXiv	2025	✓	✗	✗	~	✗	✗
Embodied Learning [305]	MIR	2025	~	✗	~	~	✗	✗
FM for Manipulation [135]	arXiv	2024	✓	~	~	~	✗	✗
Ours (Weights or Skills?)	–	2026	✓	✓	✓	✓	✓	✓

Table 1. Topical coverage of the closest related surveys versus ours. ✓ covered, ~ partial/mentioned, ✗ not covered. Columns: **weights** (VLA/weights), **code** (code-as-policy), **skill** (skill libraries), **reward** (reward synthesis), **Self-impr.** (organised by *degree of self-improvement*, incl. the §3.1.5 cell), **Economy** (the robot skill marketplace). Only this survey covers the last two axes. CSUR = ACM Computing Surveys; MIR = Machine Intelligence Research.

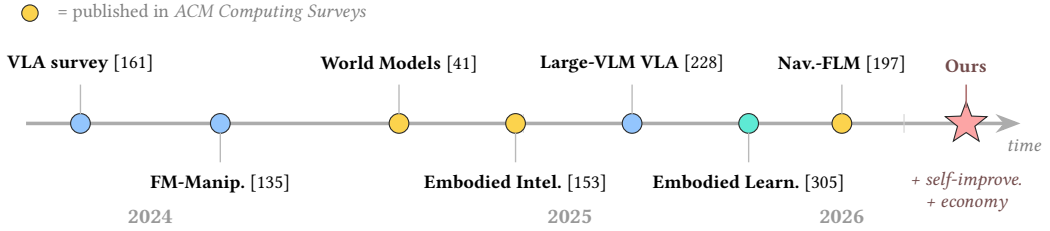


Fig. 3. The closest related surveys as a timeline (2024–2026); **amber** markers denote surveys published in ACM Computing Surveys (CSUR). Prior surveys cover VLA models, foundation models, world models, and navigation, but none organises code-as-policy by *degree of self-improvement*, the weights–skills axis of Figure 1. **Ours** (★, 2026) is organised around the self-improvement axis (§3.1.5) and the skill economy; per-topic coverage is in Table 1.

2 A TAXONOMY OF ROBOT-LEARNING TECHNIQUES

Figure 2 organises the field into six branches. The dividing question is *what ships*: frozen weights (§3.2) or executable skills (§3.1). We tag mechanisms by **Feedback**, **Search**, and **Memory** throughout.

Table 2 compares all systems in the map on what each *ships*, where it is evaluated, and whether it improves from its own experience.

Table 2. Detailed comparison of a representative 47 of the 77 systems in the taxonomy (Figure 2), grouped by family; the §3.1 and §3.4 branches are tabulated in full in Tables 4 and 6. **Ships**: **code/weights/reward/skill/policy/bench**. **Eval**: **real/sim/game**. **Self-impr.**: ✓ learns from its own experience, ~ partially, ✗ no, – n/a.

System	Year	Venue	Ships	Eval	Self-impr.
§3.1 Code-as-policy for robots					code
Code-as-Policies [144]	2023	ICRA	code	real	✗
ProgPrompt [236]	2023	ICRA	code	real	✗
VoxPoser [100]	2023	CoRL	code	real	✗
RoboCodeX [186]	2024	ICML	code	sim	✗

Table 2 (continued)

System	Year	Venue	Ships	Eval	Self-impr.
Instruct2Act [97]	2023	arXiv	code	sim	✗
RoboPro [273]	2025	arXiv	code	real	✗
RoboCoder [137]	2024	arXiv	code	sim	✗
CaP-X [58]	2026	arXiv	code	sim	✓
RoboClaw [139]	2026	arXiv	code	real	✓
ASPIRE [160]	2026	arXiv	code	real	✓
ENPIRE [272]	2026	arXiv	code	real	✓
§3.2 End-to-end VLA / generalist policies					weights
RT-1 [17]	2023	RSS	weights	real	✗
RT-2 [18]	2023	CoRL	weights	real	✗
Octo [66]	2024	RSS	weights	real	✗
OpenVLA [120]	2024	CoRL	weights	real	✗
RoboFlamingo [142]	2024	ICLR	weights	sim	✗
CogACT [138]	2024	arXiv	weights	real	✗
SpatialVLA [207]	2025	arXiv	weights	real	✗
π_0 [15]	2024	arXiv	weights	real	✗
$\pi_{0.5}$ [14]	2025	arXiv	weights	real	✗
GR00T N1 [194]	2025	arXiv	weights	real	✗
Gemini Robotics [65]	2025	arXiv	weights	real	✗
§3.3 LLM-authored rewards / curricula					reward
Eureka [163]	2024	ICLR	reward	sim	✓
DrEureka [164]	2024	RSS	reward	real	✓
Text2Reward [275]	2024	ICLR	reward	sim	~
Language-to-Rewards [288]	2023	CoRL	reward	real	~
Eurekaverse [146]	2024	CoRL	reward	sim	✓
RoboGen [256]	2024	ICML	reward	sim	✓
Auto MC-Reward [136]	2024	CVPR	reward	game	✓
§3.4 Robot skill libraries & lifelong learning					skill
Voyager [250]	2023	arXiv	skill	game	✓
LOTUS [248]	2024	ICRA	skill	sim	✓
LRLL [242]	2024	ICRA	skill	sim	✓
BOSS [297]	2023	CoRL	skill	sim	✓
SPRINT [295]	2024	ICRA	skill	sim	~
Uni-Skill [274]	2026	ICRA	skill	real	✓
SkillFlow [300]	2026	arXiv	skill	text	✓
§3.5 Sim-to-real & cross-embodiment transfer					policy
Open X-Embodiment [195]	2024	ICRA	policy	real	✗
CrossFormer [44]	2024	CoRL	policy	real	✗
RoboCat [16]	2023	TMLR	policy	real	✓
Mirage [29]	2024	RSS	policy	real	✗
§3.6 Embodied benchmarks & simulators					bench
LIBERO [149]	2023	NeurIPS	bench	sim	—
Robosuite [310]	2020	arXiv	bench	sim	—
BEHAVIOR-1K [134]	2024	arXiv	bench	sim	—
Meta-World [286]	2019	CoRL	bench	sim	—
RLBench [105]	2020	RA-L	bench	sim	—
ManiSkill2 [71]	2023	ICLR	bench	sim	—
CALVIN [178]	2022	RA-L	bench	sim	—

2.1 Survey methodology: building and partitioning the corpus

Figure 4 summarises the corpus construction in the PRISMA 2020 style [196]. The corpus was assembled in two passes over a January 2016 to July 2026 window. First, the 77 taxonomy systems were curated by *seeding and snowballing*: for each of the six branches we started from its landmark systems and chased citations forward and backward until new candidates stopped appearing. Second, the 225-work landscape corpus (Table 11) was gathered by five *structured web-search sweeps*, one per area cluster: (i) generalist and vision-language-action policies, imitation and diffusion policies; (ii) language-model planning, task-and-motion planning, reward and data synthesis; (iii) skill discovery, hierarchical reinforcement learning, and world models; (iv) manipulation, dexterity, sim-to-real, locomotion, and benchmarks; and (v) representation learning, offline reinforcement learning, navigation, and tactile sensing. Each sweep queried the open web with area-specific search strings of the form “<system or family name> robot learning arXiv” and “<area> survey/benchmark/dataset” over arXiv, the ACM Digital Library, IEEE Xplore, Google Scholar, and venue proceedings pages (CoRL, RSS, ICRA, IROS, NeurIPS, ICML, ICLR, CVPR).

Inclusion required a work to (a) concern how an embodied agent’s skill or policy is authored, learned, transferred, evaluated, or distributed, and (b) carry verifiable metadata: exact title, first author, year, and an arXiv identifier or DOI, each confirmed against the source. *Exclusion* removed pure perception, navigation, and locomotion work with no bearing on skill authoring (retained only in the landscape appendix where it anchors an area), works whose metadata could not be verified, and duplicates, detected by both bibliographic key and arXiv identifier. Every reference in this survey passed the verification step; candidates that failed it were discarded at harvest time, and we report that their count was not logged rather than estimate it. Included systems were then mapped onto the taxonomy of Figure 2 by the mechanism definitions of Table 3.

Placement: why a system enters the taxonomy or the landscape. Corpus membership and placement *within* the corpus are two separate decisions, and it is the second that produces the 77/225 split. A verified work is *taxonomised*, that is, promoted to the 77 systems of Figure 2 and scored on all six axes of Tables 2–6, only when it passes three tests. (i) **Axis-placeable**: it takes an unambiguous position on the weights-versus-skills question and, for skill-authoring systems, on the self-improvement ladder of §3.1, so that its branch is determined rather than argued. (ii) **Distinct**: it contributes a mechanism, rung, or branch exemplar not already carried by a system in the map, rather than a near-variant, ablation, or reimplementations of one. (iii) **Fully characterisable**: its primary source reports enough of what it *ships*, how it learns, where it is evaluated, and whether it self-improves to fill every comparison column without inference. A verified work that fails any one test is kept rather than dropped: it stays in the landscape appendix (Table 11) as coverage evidence, which is where precursors, same-point variants, and the supporting datasets, benchmarks, and simulators live. Two further rules keep the map legible on a single page. First, when several systems make the same taxonomic point, we taxonomise the earliest or most load-bearing landmark and catalogue the rest in the landscape. Second, pure perception, navigation, or locomotion work is taxonomised only when it directly authors or improves a manipulation skill, and otherwise anchors its area in the landscape alone. The 77/225 split is therefore a statement about analytical *role*, sharp exemplar versus representative breadth, not about a work’s quality or importance.

2.2 The corpus at a glance

Figure 5 summarises the whole corpus at a glance, and Figure 6 profiles the 225 landscape works (Table 11, Appendix A) by publication year, evaluation domain, learning signal, and embodiment; every count is the exact tally of that table. Figure 7 then charts the two poles of Figure 1 by year

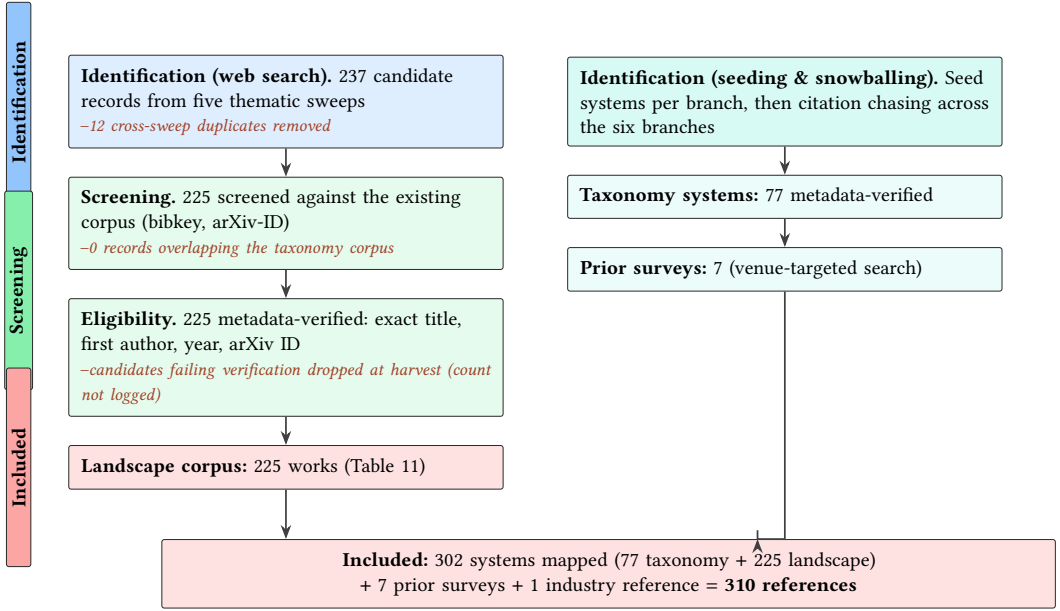


Fig. 4. **Corpus construction** in the PRISMA 2020 style [196]. The left column is the web-search stream behind the landscape corpus (Table 11), with recorded counts and inline exclusions (red) at each stage; the right column is the seed-and-snowball curation behind the weights–skills taxonomy corpus (Figure 2). Both streams feed the included set. Candidates whose metadata could not be verified were discarded at harvest time; their count was not logged, which we state rather than estimate (§2.1).

of release, showing both paradigms emerging only recently, with code-as-policy the larger and faster-growing pole.

Figures 8 and 9 then give a visual overview of the surveyed systems: the real and simulated robots they run on (Figure 8) and the empirical results they report (Figure 9); the two poles of the opening thesis figure (Figure 1) recur here across every embodiment. Every panel is a live hyperlink to its source paper and is cropped from a specific figure of it; Table 10 records the source paper and the figure number for each, so that no image is used without a traceable, clickable origin. The figures are vector graphics with the photographs embedded at full resolution, so they can be zoomed without pixelation.

3 THE TECHNIQUE FAMILIES

We now examine each of the six branches of the taxonomy (Figure 2) in turn. We begin with the code-as-policy camp (§3.1), the focus of this survey and the only family we organise by *degree of self-improvement*, and then survey the weights-based (§3.2), reward-synthesis (§3.3), skill-library (§3.4), transfer (§3.5), and benchmark (§3.6) families that surround it.

3.1 Code-as-Policy: Robots that Write their Own Skills

The organising contribution of this survey is to arrange code-as-policy methods not by application or robot embodiment, but by *how much of their own competence a system produces at run time*. Concretely, we ask five questions in sequence: does the agent (i) write control code at all, (ii) repair that code from execution feedback within a task, (iii) remember validated code across tasks, (iv) search a population of programs rather than following a single trajectory of repairs, and (v) close

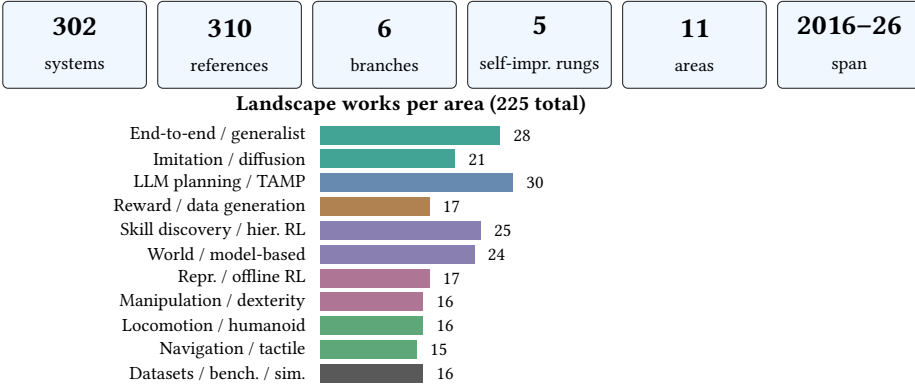


Fig. 5. **The surveyed corpus at a glance.** The survey covers 302 systems (77 placed in the taxonomy of Figure 2 plus 225 landscape works of Table 11 in Appendix A) across 6 technique branches, with the code-as-policy branch resolved into 5 self-improvement rungs, spanning 2016–2026 and totalling 310 references. The bars give the number of landscape works in each of the 11 areas, taken directly from the area tallies of Table 11. The teal and blue bars mark the survey’s two poles (Figure 1): the weights side (end-to-end and imitation-learned policies) and the skills side (code-as-policy planning); the remaining branches sit off that axis.

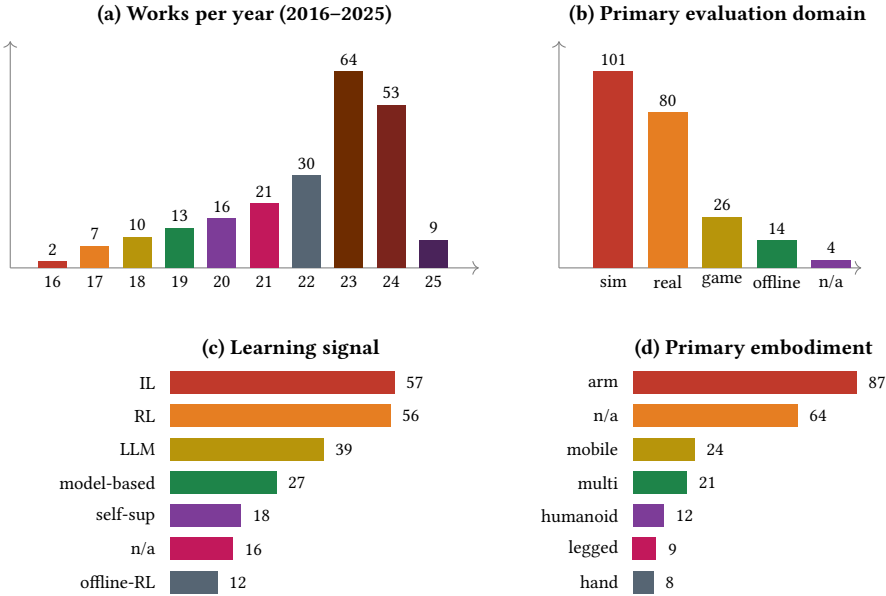


Fig. 6. **Profile of the 225 landscape works** (Table 11, Appendix A), counted directly from the taxonomy tags. (a) publication year, showing the surge in 2023–2024; (b) primary evaluation domain (simulation dominates, with real-robot a close second); (c) primary learning signal; (d) primary embodiment (“n/a” marks datasets, benchmarks, and simulators, which have none). The four profiles cut across both poles of the weights–skills axis (Figure 1), so their bars use a categorical palette that avoids the reserved pole colours (teal and blue). Counts are the exact column tallies of Table 11; they characterise the surveyed corpus, not the field’s total output.

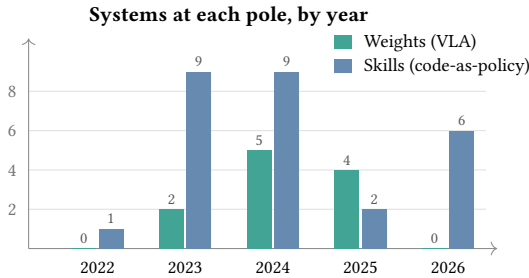


Fig. 7. **The two poles over time.** Year of first release of the taxonomy systems this survey places at each pole of Figure 1: the 11 end-to-end/VLA systems that ship *weights* (§3.2; teal) and the 27 code-as-policy systems that ship *skills* (§3.1; blue). Both paradigms are recent, and code-as-policy is the larger, faster-growing pole under this survey’s deep-dive; the counts reflect that coverage, not the field’s total output (2026 partial, through July). Years are from the bibliography; pole membership is the branch assignment of Figure 2.

Mechanism	Counts (a system has it if and only if...)	Does not count
Feedback (F)	an execution-grounded signal is read and used to revise the code <i>within</i> a task: a failure detector firing, a sensory or textual summary of what went wrong, or a verifier or reward obtained by running the program (a simulated reward qualifies when the agent uses it to revise)	a fresh human instruction; a fixed pre-trained value function that is never queried at run time; a natural-language critique that is not grounded in execution
Memory (M)	content written <i>at run time</i> is stored across task boundaries and retrieved for later tasks: a persistent code-skill library, or a retrievable episodic-text memory of past solutions	the model’s frozen pre-trained weights; a within-task scratchpad discarded at episode end; a fixed API library the agent cannot extend
Search (S)	more than one candidate program is maintained, scored by execution, and selected or mutated among: sampled-program selection, evolutionary mutation, or beam search	a single sequential chain of repairs on one candidate (that is Feedback); one-shot sampling with no selection; repeated prompting whose candidates are never compared

Table 3. Operational definitions of the three mechanisms used to place a system on the code-as-policy ladder (§3.1). These resolve the ambiguous boundary cases: a failure detector or a grounded critique counts as Feedback, trained weights do not count as Memory, and repeated prompting counts as Search only when its candidates are compared and selected.

all of these into one open-ended loop? Each “yes” is a rung on the ladder shown in the §3.1 branch of Figure 2, and the population thins sharply toward the top: of the thirteen zero-shot systems, only a handful ever reach the combined feedback, memory, and search regime. The purpose of this arrangement is to name the progression and to make explicit how sparsely the top rung is populated. We use the operational definitions of Table 3 throughout: a system exhibits feedback, memory, or search only under the conditions stated there.

3.1.1 Zero-shot synthesis. Membership condition: a system is on this rung if and only if it emits executable control code but closes no execution loop back to code generation, so feedback, memory, and search are all absent. The foundational move is to treat the language model as a *program synthesiser* over a fixed library of perception and control primitives. *Code-as-Policies* [144] showed

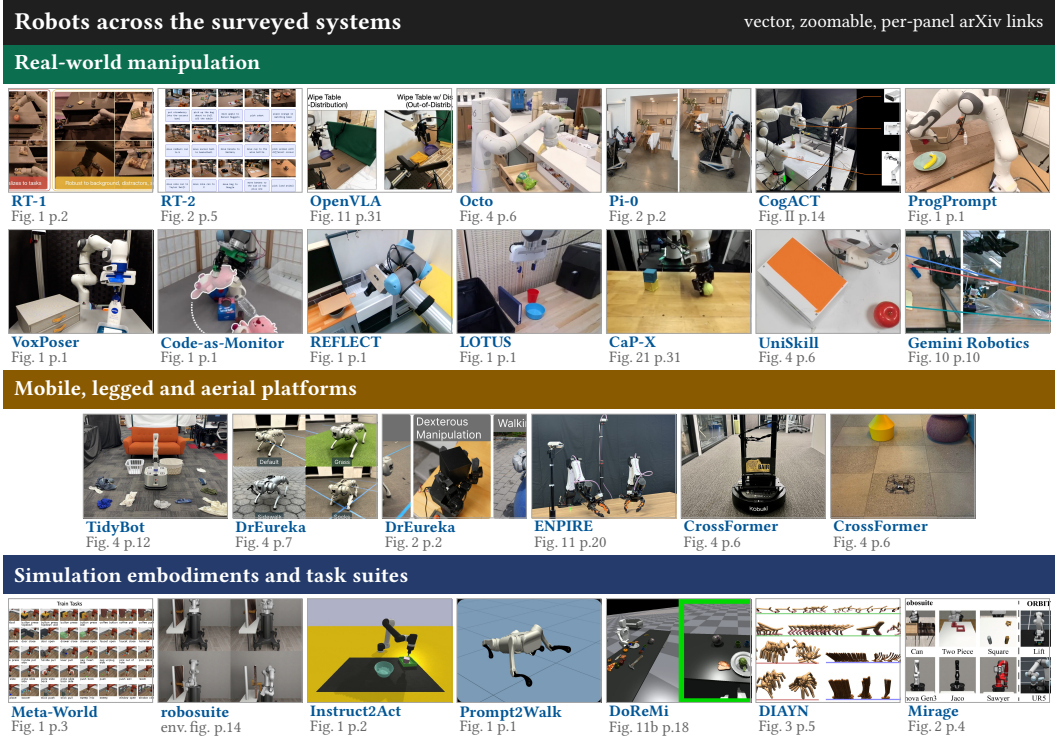


Fig. 8. **Robots across the surveyed systems.** Representative embodiments from the corpus, grouped into real-world manipulators, mobile/legged/aerial platforms, and simulation embodiments. Systems from both poles of the weights–skills axis (Figure 1) appear here, grouped by embodiment rather than by pole because the two paradigms share the same hardware. Each panel is a live hyperlink to the source paper on arXiv and is annotated with the exact figure number it was cropped from (also listed in Table 10); the figure is vector with photographs embedded at full resolution, so it can be zoomed without pixelation. Only images that depict a robot are shown. Images © their respective authors, reproduced for scholarly review.

that an LLM prompted with an API and a natural-language instruction can emit executable Python that composes those primitives, recursively defining undefined functions and parameterising control with arithmetic and feedback logic. *ProgPrompt* [236] situates the generated program in the scene by exposing available actions and objects as importable symbols, while *VoxPoser* [100] has the model write code that *composes 3D value maps* for a motion planner rather than calling fixed skills, loosening the dependence on a hand-authored API. Subsequent work broadened the input modality and task horizon; *RoboCodeX* [186] conditions on multimodal observations, *Instruct2Act* [97] maps instructions to perception–action code, and video- and demonstration-conditioned variants such as *RoboPro* [273], *Demo2Code* [251] and *Text2Motion* [147] recover programs from richer context. *Statler* [285] maintains an explicit world state across steps, and *TidyBot* [265] and *ChatGPT for Robotics* [244] demonstrate personalisation and prompt design for real hardware. What unites this rung, and limits it, is that the program is written *once*: there is no channel from execution back to the synthesiser, so a plan that fails at run time simply fails.

3.1.2 Closed-loop self-repair. Membership condition: feedback is present but memory and search are absent; the agent revises code within a task from an execution-grounded signal yet retains

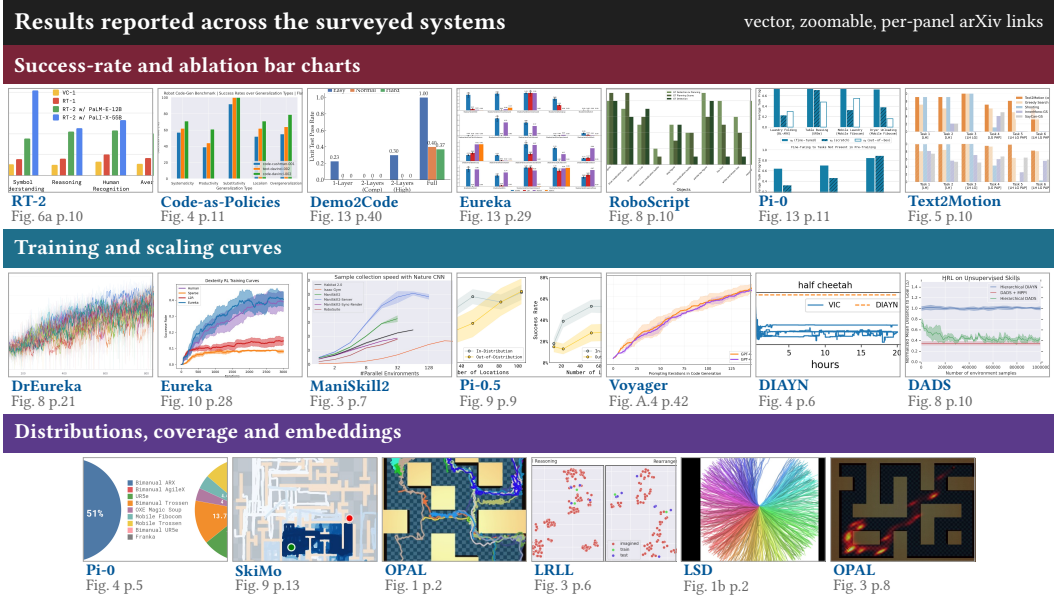


Fig. 9. **Results reported across the surveyed systems.** Representative results figures from the corpus, grouped into success-rate/ablation bars, training and scaling curves, and distribution, coverage and embedding plots. The results come from systems on both sides of the weights-skills axis (Figure 1), grouped here by type of evidence. Each panel is a live hyperlink to the source paper on arXiv and is annotated with the exact figure number it was cropped from (also in Table 10). Panels convey the shape and variety of empirical evidence, not absolute comparisons across incommensurable protocols. Images © their respective authors, reproduced for scholarly review.

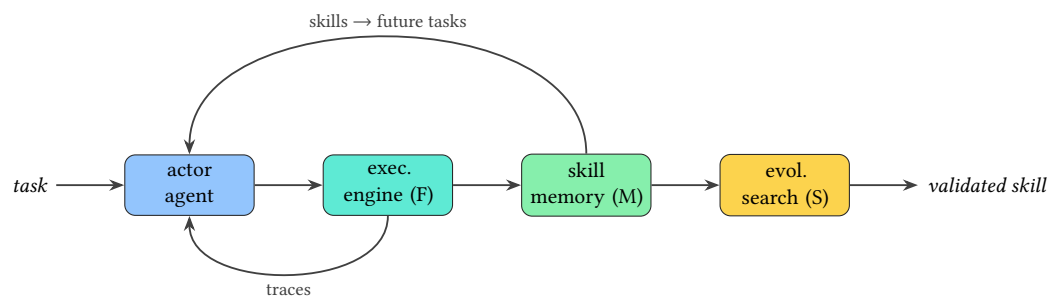


Fig. 10. The full self-improving loop (§3.1.5): feedback (F) + memory (M) + search (S) combine into one open-ended loop (occupied by ASPIRE, ENPIRE, and RoboClaw). VLAs (§3.2) have no such loop.

nothing across tasks and never compares a population of candidates. The second rung adds a within-task feedback loop. *Inner Monologue* [102] feeds success detectors, scene descriptions and human feedback back into the model as language, letting it replan when a step fails. *DoReMi* [73] detects misalignments between plan and execution and triggers recovery, and *REFLECT* [159] builds a hierarchical, multimodal summary of past interaction that an LLM queries to explain a failure and propose a correction. More recent systems push the monitor into the perception loop; *Code-as-Monitor* [306] compiles spatio-temporal constraints into vision-language checks, and *AHA*

[48] trains a model to reason explicitly about failure modes. The novelty of this rung is the closed loop, but it is a *short* loop: corrections are discarded at task boundaries, so nothing accumulates.

3.1.3 Skill-library accumulation. *Membership condition:* memory is present; validated code is written to a cross-task store and retrieved later, independently of whether within-task feedback is used. The third rung makes the loop persistent by writing validated code into a growing library that later tasks retrieve and compose. Because this “memory” axis is a full technique family in its own right, we catalogue its systems once, in §3.4 (§3.4.3); *Voyager* [250] is the canonical example, curating an ever-growing library of verified skill programs with when-to-apply guards, while related methods distil language corrections into retrievable knowledge for future tasks [294]. The key distinction from §3.1.2 is temporal scope: competence compounds *across* tasks rather than being rebuilt each time.

3.1.4 Evolutionary program search. *Membership condition:* search is present; the agent maintains and selects among more than one candidate program rather than repairing a single one. The fourth rung replaces a single chain of repairs with population-based search over programs. Rather than iteratively debugging one candidate, the agent maintains several, scores them by execution, and mutates the best. *CaP-X* [58] provides an interactive gym and benchmark for exactly this style of coding agent, and *RoboEvolve* [26] couples a planner and a learned simulator into a co-evolutionary loop that mines near-miss failures to stabilise search. The novelty is the shift from *repair* to *search*: exploration is explicit and parallel, which trades compute for a better chance of escaping local failures.

3.1.5 The full self-improving loop. *Membership condition:* all three mechanisms are present in one closed loop, so the rung is the conjunction feedback *and* memory *and* search (Figure 10). Only a few very recent systems satisfy it. *ASPIRE* [160] pairs a robot execution engine that emits per-primitive multimodal traces with a skill library of validated code and a population search over candidate programs; *ENPIRE* [272] closes an analogous loop on physical hardware through automatic reset, parallel rollouts, and log-driven revision; and *RoboClaw* [139] unifies data collection, policy learning, and execution under a single controller with self-resetting loops. That this cell is so sparsely populated, rather than any particular system within it, is the structural observation the survey is organised to make.

Table 4 makes the funnel explicit, tabulating each system’s use of feedback (F), memory (M), and search (S).

Table 4. The 30 systems on the code-as-policy self-improvement ladder: the 27 of the §3.1 branch of Figure 2 plus the three §3.4.3 skill-library systems (*Voyager*, *LRLl*, *Uni-Skill*) that also occupy its memory rung, on the three self-improvement mechanisms, **Feedback**, **Memory**, **Search**, which are determined by the rung. ✓ present, ✗ absent. **Self-impr.:** ✓ full loop, ~ partial, ✗ none.

System	Year	F	M	S	Self-impr.
§3.1.1 Zero-shot synthesis					F ✗ M ✗ S ✗
Code-as-Policies [144]	2023	✗	✗	✗	✗
ProgPrompt [236]	2023	✗	✗	✗	✗
VoxPoser [100]	2023	✗	✗	✗	✗
Instruct2Act [97]	2023	✗	✗	✗	✗
ChatGPT for Robotics [244]	2023	✗	✗	✗	✗
TidyBot [265]	2023	✗	✗	✗	✗
RoboCodeX [186]	2024	✗	✗	✗	✗
RoboScript [27]	2024	✗	✗	✗	✗
Text2Motion [147]	2023	✗	✗	✗	✗

Table 4 (continued)

System	Year	F	M	S	Self-impr.
Demo2Code [251]	2023	✗	✗	✗	✗
Statler [285]	2024	✗	✗	✗	✗
Prompt2Walk [257]	2024	✗	✗	✗	✗
RoboPro [273]	2025	✗	✗	✗	✗
§3.1.2 Closed-loop self-repair				F ✓	M ✗ S ✗
Inner Monologue [102]	2022	✓	✗	✗	✗
DoReMi [73]	2024	✓	✗	✗	✗
REFLECT [159]	2023	✓	✗	✗	✗
Code-as-Monitor [306]	2025	✓	✗	✗	✗
AHA [48]	2024	✓	✗	✗	✗
Introspective Planning [145]	2024	✓	✗	✗	✗
§3.1.3 Skill-library accumulation				F ✗	M ✓ S ✗
Voyager [250]	2023	✗	✓	✗	~
LRLl [242]	2024	✗	✓	✗	~
RoboCoder [137]	2024	✗	✓	✗	~
Uni-Skill [274]	2026	✗	✓	✗	~
DROC [294]	2024	✓	✓	✗	~
§3.1.4 Evolutionary program search				F ✓	M ✗ S ✓
CaP-X [58]	2026	✓	✗	✓	~
RoboEvolve [26]	2026	✓	✗	✓	~
Code Evolution (GEAR) [72]	2026	✓	✗	✓	~
§3.1.5 Full self-improving loop				F ✓	M ✓ S ✓
ASPIRE [160]	2026	✓	✓	✓	✓
ENPIRE [272]	2026	✓	✓	✓	✓
RoboClaw [139]	2026	✓	✓	✓	✓

3.2 End-to-End Vision-Language-Action Models

The dominant alternative to writing code is to regress actions directly from pixels and language with a single large network, so that competence lives in *frozen weights* rather than in an inspectable program. Architecturally these vision-language-action (VLA) models share a backbone–action-head decomposition: a vision-language backbone encodes the scene and instruction, and a pluggable action head maps that representation to motor commands [15, 120].

The lineage begins with *RT-1* [17], a transformer trained on large-scale real robot data that discretises continuous actions into categorical bins. *RT-2* [18] then co-fine-tunes an Internet-pretrained vision-language model on robot trajectories, emitting actions as text tokens and inheriting semantic generalisation from web data. Open reproductions followed: *Octo* [66] and *OpenVLA* [120], the latter a 7B model with a DINOv2/SigLIP vision stack, are trained on the pooled Open X-Embodiment corpus (§3.5), and *RoboFlamingo* [142] adapts a vision-language model into an imitation policy.

A second design axis is the *action representation*. Discrete token heads (RT-1/RT-2) are simple but coarse; recent systems favour continuous heads. π_0 [15] attaches a *flow-matching* head to a PaLI-Gemma backbone for smooth, high-frequency bimanual control, and $\pi_{0.5}$ [14] extends it toward open-world generalisation; *CogACT* [138] separates a cognition module from a diffusion-transformer action module, and *SpatialVLA* [207] injects 3D spatial encodings. Flow- and diffusion-based heads trade a few extra parameters for far fewer inference steps and better high-degree-of-freedom precision.

At the largest scale, foundation-model efforts target whole platforms: *GR00T N1* [194] pairs a slow System-2 vision-language reasoner with a fast System-1 diffusion transformer that produces

motor actions at high rate for humanoids, and *Gemini Robotics* [65] builds a VLA together with an embodied-reasoning model on the Gemini backbone. Across all of these, the defining property, and the reason we treat them as the *foil* for this survey, is that there is *no loop and no code*: competence is acquired by scaling data and parameters, generalises by interpolation rather than search, and cannot be edited, audited, or recombined after training.

3.3 Reward and Curriculum Synthesis

A third family keeps the language model in the role of programmer but points it at a different artefact: instead of policy code, it writes the *reward*, *environment*, or *curriculum* that a reinforcement-learning loop then compiles into a policy. A feedback loop exists, the model reads back training statistics and revises, but what ships is still weights, which is why we separate this family from the self-improving code camp (§3.1).

Eureka [163] is the canonical instance: given unmodified environment source code and a task description, a coding LLM zero-shot generates an executable reward function, then improves it by *evolutionary search* guided by *reward reflection*, a textual summary of training statistics, exceeding expert human rewards on 83% of a 29-task IsaacGym suite. *DrEureka* [164] extends the idea to sim-to-real by also synthesising domain-randomisation ranges, achieving zero-shot real-world locomotion. *Text2Reward* [275] generates dense reward programs grounded in a compact environment representation and refines them from human feedback, matching or beating expert-written rewards on most of a seventeen-task suite, while *Language to Rewards* [288] uses the reward function as the interface between an LLM and a MuJoCo model-predictive controller.

Beyond the reward itself, the same generative recipe is applied to the *training environment*. *Eurekaverse* [146] has an LLM propose a progressively harder curriculum of environments, learning parkour skills that transfer to a real robot, and *RoboGen* [256] runs a propose-generate-learn cycle that fabricates tasks, scenes, and supervision in simulation to unlock effectively unlimited training data; *Auto MC-Reward* [136] automates dense rewards for open-world Minecraft. The through-line is that language becomes a bridge from high-level intent to low-level control *via* an RL optimiser, rather than producing an executable, reusable skill directly.

3.4 Skill Libraries: How a Repertoire is Built

Where §3.1 asked *how* a code skill improves, this family asks the prior question of *how a repertoire of skills is built at all*. The §3.4 branch of Figure 2 organises these systems along one axis, skills learned as latent reinforcement-learning policies versus skills stored as retrievable code, and only the code flavour self-improves without gradients, which links this family back to §3.1.

Before surveying these systems we note that the word “skill” is badly overloaded. Table 5 separates five senses that the literature routinely conflates, from a latent policy to a market product, and records which properties each sense affords.

3.4.1 Unsupervised and latent skill discovery. The oldest lineage learns a diverse set of skills with *no task reward and no code*, by maximising an information-theoretic objective. *DIAYN* [50] maximises the mutual information between a latent skill code and the visited states, yielding distinguishable behaviours; *DADS* [229] makes the discovered skills dynamics-aware and hence usable for model-based control; and *LSD* [200], *CIC* [129], and *METRA* [202] push toward more dynamic, far-reaching, and scalable skills through Lipschitz constraints, contrastive objectives, and metric-aware abstractions respectively.

3.4.2 Skill-based and hierarchical reinforcement learning. A second lineage extracts a continuous *skill space* from offline data and then plans or learns within it. *SPiRL* [206] learns a skill embedding together with a skill prior that accelerates downstream RL; *OPAL* [6] discovers temporally extended

Sense of “skill”	Representation	Examples	Inspect.	Adapt.	Compos.	Distrib.
latent policy	latent-conditioned network $\pi(a s, z)$	DIAYN, DADS, ME-TRA	✗	~	✗	✗
option / primitive	temporally-extended sub-policy with a skill prior	SPiRL, OPAL, SkiMo	~	~	✓	✗
code	executable program over an API	Code-as-Policies, Voyager	✓	✓	✓	~
robot app	packaged, one-tap deployable behaviour	UniStore motion packages	✗	✗	✗	✓
market product	a listed, versioned, priced commodity	UniStore listings	✗	✗	~	✓

Table 5. The word “skill” spans at least five senses that the literature frequently conflates. Columns record whether a skill is human-**inspectable**, **adaptable** after creation, symbolically **composable**, and **distributable** across robots (✓ yes, ~ partial, ✗ no). Only the *code* sense is inspectable, adaptable, and composable at once; only the app and market senses are readily distributable, which is the mismatch the skill economy (§4) must close.

primitives for offline RL; *PARROT* [235] learns an invertible behavioural prior; and *SkiMo* [231] learns a skill dynamics model so that planning happens directly in skill space.

3.4.3 LLM and code skill libraries. The code flavour, where the self-improving methods of §3.1 live, stores skills as retrievable programs. *Voyager* [250] curates an ever-growing library of verified code skills in Minecraft; *LOTUS* [248] performs continual imitation learning through unsupervised skill discovery; *LRL* [242] bootstraps a composable robot library with language models; and *BOSS* [297] and *SPRINT* [295] grow skill repertoires by LLM-guided bootstrapping and instruction relabelling. The most recent entries close the loop between video and code: *Uni-Skill* [274] builds a self-evolving skill repository from unstructured robot video, and *SkillFlow* [300] benchmarks lifelong discovery, patching, and reuse of a skill library.

3.4.4 Open-world LLM-agent libraries. Finally, embodied LLM agents accumulate skills from open-ended experience. *GITM* [309] and *JARVIS-1* [259] pair a planner with text or multimodal memory in Minecraft; *Expel* [301] extracts reusable insights from a stream of trials; *Optimus-1* [143] adds a hybrid multimodal memory for long-horizon tasks; and *Odyssey* [157] equips an agent with an open-world library of primitive and compositional skills. These systems foreshadow the marketplace dynamics of §4: a library that grows with use.

Table 6 contrasts the four sub-families on skill form, whether they are gradient-trained or LLM-authored, whether they keep a persistent library, and their domain.

Table 6. The 21 skill-building systems of the §3.4 branch of Figure 2, by sub-family. **Form:** latent-RL/skill-space/code-lib/agent-lib. **Grad.:** gradient-trained (✓) vs LLM-authored (✗). **Lib.:** persistent retrievable library. **Domain:** sim/game/real/text.

System	Year	Form	Grad.	Lib.	Domain
§3.4.1 Unsupervised / latent discovery (RL)					
DIAYN [50]	2019	latent-RL	✓	✗	sim
DADS [229]	2020	latent-RL	✓	✗	sim
LSD [200]	2022	latent-RL	✓	✗	sim
CIC [129]	2022	latent-RL	✓	✗	sim

Table 6 (continued)

System	Year	Form	Grad.	Lib.	Domain
METRA [202]	2024	latent-RL	✓	✗	sim
§3.4.2 Skill-based / hierarchical RL					
SPiRL [206]	2020	skill-space	✓	✗	sim
OPAL [6]	2021	skill-space	✓	✗	sim
PARROT [235]	2021	skill-space	✓	✗	sim
SkiMo [231]	2022	skill-space	✓	✗	sim
§3.4.3 LLM / code skill libraries					
Voyager [250]	2023	code-lib	✗	✓	game
LOTUS [248]	2024	code-lib	✗	✓	sim
LRL [242]	2024	code-lib	✗	✓	sim
BOSS [297]	2023	code-lib	✗	✓	sim
SPRINT [295]	2024	code-lib	✗	✓	sim
Uni-Skill [274]	2026	code-lib	✗	✓	real
SkillFlow [300]	2026	code-lib	✗	✓	text
§3.4.4 Open-world LLM-agent libraries					
GITM [309]	2023	agent-lib	✗	✓	game
JARVIS-1 [259]	2023	agent-lib	✗	✓	game
ExpeL [301]	2024	agent-lib	✗	✓	text
Optimus-1 [143]	2024	agent-lib	✗	✓	game
Odyssey [157]	2024	agent-lib	✗	✓	game

3.5 Sim-to-Real and Cross-Embodiment Transfer

A skill is only useful if it moves across the simulation-to-reality gap and across robot bodies. The enabling artefact is scale: *Open X-Embodiment* [195] pools sixty datasets from twenty-one institutions into over a million trajectories spanning twenty-two embodiments and hundreds of skills, and shows that RT-X models trained on the mixture exhibit positive transfer and emergent capabilities across platforms.

Given such data, a single network can control very different bodies. *CrossFormer* [44] trains one transformer on 900K trajectories across twenty embodiments, arms, wheeled robots, quadrotors, and quadrupeds, *without* manually aligning observation or action spaces, and matches specialist policies tailored to each robot. Other work attacks transfer at test time rather than through data: *Mirage* [29] achieves *zero-shot* cross-embodiment transfer by “cross-painting”, masking the target robot and inpainting the source robot at the same pose so the policy sees a familiar body, and bridging the control gap with forward dynamics.

The most suggestive point for this survey is *RoboCat* [16], a goal-conditioned multi-embodiment agent that adapts to a new task from as few as a hundred demonstrations and then *generates its own data* for the next training round, forming a rudimentary self-improvement loop of the kind §3.1 makes explicit. This is also the marketplace question of §4: a skill advertised for the G1, H1, B2, and Go2 platforms must survive exactly the cross-embodiment gap these methods study.

3.6 Benchmarks and Evaluation

Progress in the families above is measured on a shared set of simulated benchmarks, most of which report a single-number success rate. *LIBERO* [149] targets lifelong manipulation and knowledge transfer across task suites; *Meta-World* [286] offers fifty manipulation tasks for multi-task and meta-reinforcement learning; *RLBench* [105] provides a hundred vision-based tasks with demonstrations; and *ManiSkill2* [71] adds GPU-parallel simulation over twenty task families. *Robosuite* [310] is the modular MuJoCo framework underlying many of these, and *CALVIN* [178] evaluates long-horizon,

Dimension	Full loop [139, 160, 272] <i>code+skills</i>	π_0 [15] <i>VLA weights</i>	Code-as-Pol. [144] <i>code, no mem.</i>	Eureka [163] <i>reward</i>
What it outputs	skill code+lib	action chunks	policy code	reward code
How it improves	F+M+S loop	gradient pretrain	– one-shot	RL+reflection
Gradient training?	✗	✓	✗	✓
Persistent skill library?	✓	✗	✗	✗
Failure feedback signal	exec. traces	–	–	train. stats
Search strategy	evolutionary	–	–	–
Real-robot validation	✓	✓	✓	✓
Continual / open-ended?	✓	✗	✗	✗
Beyond fixed APIs?	✓	✓	✗	–
Interpretable / editable?	✓	✗	✓	✓

Table 7. Capability matrix, one representative per camp. Only the full-loop cell (§3.1.5; e.g. ASPIRE, ENPIRE, RoboClaw) combines a persistent skill library with feedback- and search-driven self-improvement; each other camp lacks at least one axis. ✓ = yes, ✗ = no, – = n/a.

language-conditioned control by the average number of sub-tasks completed in a chain. At the high-realism end, *BEHAVIOR-1K* [134] specifies a thousand everyday household activities in a photorealistic simulator, scoring both success and efficiency.

Table 7 contrasts one exemplar per camp on the axes that separate them. The recurring gap, and a direct motivation for the open problems of §4, is that these suites measure *one-shot competence*: they report whether a policy succeeds, not whether it *improves with experience*. No standard benchmark yet plots held-out success as a function of accumulated interaction, which is exactly the quantity the self-improvement ladder of §3.1 is designed to raise.

3.7 Synthesis: what each family cannot do

Table 8 compares every family on six axes: data needs, task horizon, transfer, interpretability, safety, and characteristic failure mode. Reading it column-wise profiles a family; reading it row-wise shows that every axis has at least one weak family, and no family is favourable on all six, which is the tension the skill economy (§4) inherits.

4 THE SKILL ECONOMY: OPEN PROBLEMS

The arrival of a commercial marketplace for robot skills (Unitree’s UniStore ships one-tap, cross-model motion packages [243]) turns several research questions from hypothetical into pressing. We frame them as the open agenda of this survey.

The static-to-adaptive gap. Every skill shipped today is, in effect, replayed: a motion package or a fixed program is installed and executed without perceiving whether *this* kitchen, gripper, or object differs from the one it was authored for. The techniques of §3.1, within-task repair, persistent memory, and search, are precisely what would let a downloaded skill *adapt* on the target robot. Closing this gap is the difference between a store of animations and a store of capabilities, and it is the direct application of the §3.1.5 loop to a distribution setting where the base skill is given rather than discovered.

Cross-embodiment portability. UniStore advertises a single skill running across the G1 and H1 humanoids and the B2/Go2 quadrupeds; bodies with different morphologies, action spaces, and

	End-to-end VLA (§3.2)	Code-as-policy (§3.1)	Reward synthesis (§3.3)	RL skill discovery (§3.4)	Market skills (§4)
Data needs	✗ web-scale teleop corpora: 1M+ episodes, 22 embodiments [195]	✓ few-shot prompts over a pretrained LLM [144]	~ massive <i>simulated</i> interaction, no demonstrations [163]	✗ millions of environment steps per embodiment [50]	✓ none at deployment; vendor-recorded playback [243]
Task horizon	~ short reactive chunks; long tasks by chaining [18]	✓ long: programs compose perception and control primitives [100]	✗ one low-level skill per synthesised reward [163]	✗ temporally-extended primitives, not full tasks [202]	~ fixed multi-step routines; no branching
Transfer	✓ cross-task and cross-embodiment via co-training [195]	~ API-level portability, but bound to the perception stack	~ sim-to-real via randomisation search [164]	~ latent skills seed downstream RL [206]	✗ certified robot models only; no adaptation [243]
Interpretability	✗ opaque weights; behaviour only observable	✓ readable, editable, diffable programs	~ readable reward code; opaque trained policy	✗ latent skill codes, symbolically uninspectable	✗ sealed vendor package
Safety	✗ emergent behaviour is hard to verify pre-deployment	✓ inspectable before execution; constrained to vetted APIs	✗ reward hacking; unintended optima [163]	✗ unconstrained exploration unsafe on hardware	~ vendor certification, but no provenance record
Failure mode	silent misgrounding under distribution shift	brittle when perception mislabels the scene	specification gaming that passes in sim only	skill collapse into degenerate behaviours	context mismatch with no recovery path

Table 8. **Family comparison matrix:** the five technique families on six axes (data needs, task horizon, transfer, interpretability, safety, characteristic failure mode). Marks grade each family on the axis (✓ favourable, ~ mixed, ✗ weak); evidence anchors include the Open X-Embodiment corpus [195] for data and transfer and lifelong suites such as LIBERO [149] for horizon and reuse. No family is favourable on all six axes, which is the tension the skill economy (§4) inherits.

dynamics. This is the cross-embodiment transfer problem (§3.5) at deployment scale: without a shared action interface [195] or explicit embodiment adaptation, a package tuned for one platform has no guarantee of correctness on another. Formal portability, what must hold for a skill to be certified on a new body, remains open.

Provenance and trust. A marketplace invites third-party, crowdsourced uploads, including raw motion-capture data and code from unknown authors. Robot skills act on the physical world, so provenance (who authored a skill, from what data, validated how) becomes a safety property, not merely a metadata nicety. There is no accepted standard for signing, attesting, or auditing a robot skill’s origin.

Safety verification. Vendors promise that “every skill is scanned, tested, and verified,” but for a code-as-policy skill this is undecidable in general and hard in practice: the skill’s effect depends on the scene it meets. Practical questions include which pre-conditions a skill must declare, what sandboxed or simulated screening it must pass, and how run-time monitors (cf. *Code-as-Monitor* [306]) can bound behaviour on unseen inputs.

Skill composition. Downloading two skills should ideally yield a third: “make coffee” plus “clear the table” composed into a morning routine. Composition requires shared interfaces and a calculus over pre-/post-conditions; today’s motion packages are opaque and do not compose, and code skills compose only within a single authoring system.

Portability standards and skill ontologies. Reuse at scale needs a shared vocabulary: declared pre-conditions, expected effects, and embodiment-capability profiles, with relations such as *requires*,

provides, and *substitutable-by*. Prior work on skill ontologies and hardware-level reusability points the way, but no cross-vendor standard exists, which is what a genuine ecosystem would require.

Local versus shared libraries. Finally, the field must reconcile two notions of “skill library” that this survey has kept distinct: the *self-built* library a robot grows from its own experience (§3.1) and the *shared* library a robot downloads from a marketplace (UniStore). The interesting regime is the hybrid, an agent that both contributes to and draws from a commons, and its incentives, quality control, and feedback dynamics are entirely unstudied.

5 LIMITATIONS OF THIS SURVEY

We state the boundaries of the survey explicitly. First, our organising axis, *degree of self-improvement*, is deliberately code-centric; it foregrounds systems that emit and revise programs and treats weight-based policies (§3.2) and reinforcement-learning skill discovery (§3.4) as context rather than as the object of the taxonomy. A survey centred on representation learning or on real-world data collection would draw the map differently. Second, the frontier we emphasise (§3.1.5) is populated by very recent, in some cases concurrent, systems; benchmark numbers across them are not directly comparable, and we therefore report capabilities qualitatively rather than ranking methods. Third, the field moves quickly: several 2025–2026 systems we cite were released while this survey was being written, and the coverage should be read as a snapshot. Finally, the “skill economy” framing (§4) is grounded in an emerging commercial trend (one vendor’s marketplace at the time of writing); we treat it as motivation for open problems, not as a mature body of results.

6 FUTURE DIRECTIONS

To make the agenda actionable rather than aspirational, we anchor it on four measurable quantities, defined with concrete testbeds in Table 9: the *success-vs-interactions curve* (does competence rise with autonomous experience?), the *skill-library reuse rate* (does stored competence compound?), the *cross-embodiment transfer drop* (how much is lost when a skill changes bodies?), and the *provenance check* (can a distributed skill’s origin and test record be verified?). The first two instantiate naturally on LIBERO-style lifelong task streams [149]; the third on held-out-embodiment splits of Open X-Embodiment [195]; the fourth on marketplace listings [243]. Figure 11 sketches the target shape each of these measurements should take.

Three directions follow directly from the map, each now with its success measure. **(1) Standardised evaluation of self-improvement.** The systems on the top rung (§3.1.5) each report improvement on their own splits; the field lacks a shared protocol that plots the success-vs-interactions curve and the skill-library reuse rate under a common interaction budget. A benchmark that publishes both curves per system, in the way LIBERO standardised lifelong manipulation [149], would make “self-improvement” a reported number rather than a claim. **(2) Adaptation as a first-class marketplace primitive.** The open problems of §4 suggest a research programme in which a downloaded skill is not a frozen artefact but a starting point that the §3.1 loop specialises to the local robot and environment; its success measures are the transfer drop before versus after on-device adaptation, and the provenance check on what was downloaded. **(3) Bridging the RL and code flavours of skills.** §3.4 shows that skills are learned either as latent RL policies or as retrievable code; a unifying interface, code that wraps, calls, and refines learned policies, and learned policies distilled back into inspectable code, would let the self-improvement machinery of §3.1 operate over both, and its progress is legible in the reuse rate crossing family boundaries. We see the combination of a standard self-improvement benchmark with an adaptation-centric marketplace as the most likely path from today’s static skill stores to genuinely capable, compounding robots.

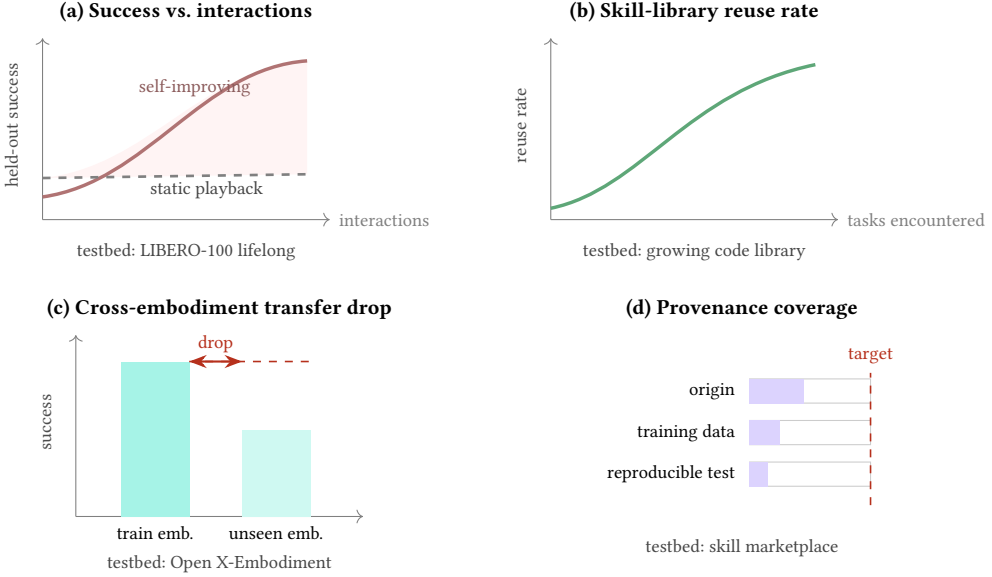


Fig. 11. **An actionable evaluation protocol for the future agenda** (§6; Table 9). The panels are *schematic*: they show the axes and the *target* shape of each measurement, not measured results. (a) held-out success should rise with autonomous interaction for a self-improving system while a static playback skill stays flat, on LIBERO-style lifelong streams [149]; (b) a growing skill library should raise its reuse rate; (c) the drop in success from a training embodiment to an unseen one quantifies transfer, on Open X-Embodiment-style splits [195]; (d) each distributed skill should carry a verifiable origin, training-data statement, and reproducible test, a coverage today far below target for marketplace skills [243].

7 CONCLUSION

Robot learning is organising itself around a single question: should competence be shipped as *weights* or as *skills*? This survey took that question as its axis. We mapped the field into six branches (§2), positioned end-to-end vision-language- action models as the weights-based foil (§3.2), and gave the code-as-policy camp its own deep-dive (§3.1) organised by *degree of self-improvement*: from one-shot program synthesis, through within-task repair, persistent skill memory, and population search, to the open-ended loop that combines all three. That top cell, feedback + memory + search, is the least populated and least surveyed region of the field, and it is where systems such as ASPIRE, ENPIRE, and RoboClaw now sit. We complemented this with a map of how skill *repertoires* are built (§3.4), showing that skills come in an RL flavour and a code flavour of which only the latter self-improves without gradients, and we connected the academic taxonomy to the nascent skill economy (§4), where commercial marketplaces already distribute skills but ship only static playback. The gap between that static distribution and genuine on-device adaptation is, we argue, the defining research opportunity for physical AI, and the self-improvement ladder of §3.1 offers a structured path toward closing it.

REFERENCES

- [1] Joshua Achiam, Harrison Edwards, Dario Amodei, and Pieter Abbeel. 2018. Variational Option Discovery Algorithms. *arXiv preprint arXiv:1807.10299* (2018).
- [2] Ademi Adeniji, Amber Xie, Carmelo Sferrazza, Younggyo Seo, Stephen James, and Pieter Abbeel. 2023. Language Reward Modulation for Pretraining Reinforcement Learning. *arXiv preprint arXiv:2308.12270* (2023).

Metric	Definition	Instantiation (testbed)	Family stressed
Success-vs-interactions curve	held-out success rate as a function of accumulated autonomous interactions; report the curve and its area, not a single endpoint	lifelong task streams in the style of LIBERO-100 [149], fixed interaction budget	§3.1 full loop; §3.2 fine-tuning
Skill-library reuse rate	fraction of new-task solutions that invoke at least one previously stored skill, and mean invocations per stored skill	growing code libraries [250] evaluated across LIBERO-style task suites [149]	§3.4 libraries; §3.1 memory rung
Cross-embodiment transfer drop	difference in success between the training embodiment and an unseen one, at matched task and budget	held-out-embodiment splits of Open X-Embodiment [195] (22 embodiments)	§3.2; §3.5
Provenance check	fraction of distributed skills carrying a verifiable origin, training-data statement, and reproducible test record	audits of marketplace listings [243] against a declared provenance schema	market skills (§4)

Table 9. **An actionable protocol for the future agenda:** four measurable quantities, their definitions, and concrete testbeds. The first two instantiate LIBERO-style lifelong learning [149]; the third instantiates Open X-Embodiment-style multi-robot transfer [195]; the fourth targets the skill marketplaces of §4.

- [3] Ananye Agarwal, Ashish Kumar, Jitendra Malik, and Deepak Pathak. 2022. Legged Locomotion in Challenging Terrains using Egocentric Vision. In *Conference on Robot Learning (CoRL)*. arXiv:2211.07638.
- [4] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, et al. 2022. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. *arXiv preprint arXiv:2204.01691* (2022).
- [5] Anurag Ajay et al. 2023. Is Conditional Generative Modeling all you need for Decision-Making?. In *International Conference on Learning Representations (ICLR)*. arXiv:2211.15657.
- [6] Anurag Ajay, Aviral Kumar, Pulkit Agrawal, Sergey Levine, and Ofir Nachum. 2021. OPAL: Offline Primitive Discovery for Accelerating Offline Reinforcement Learning. In *International Conference on Learning Representations (ICLR)*. arXiv:2010.13611.
- [7] Ilge Akkaya et al. 2019. Solving Rubik’s Cube with a Robot Hand. *arXiv preprint arXiv:1910.07113* (2019).
- [8] Eloi Alonso, Adam Jelley, Vincent Micheli, Anssi Kanervisto, Amos Storkey, Tim Pearce, and François Fleuret. 2024. Diffusion for World Modeling: Visual Details Matter in Atari. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2405.12399.
- [9] Marcin Andrychowicz et al. 2018. Learning Dexterous In-Hand Manipulation. *arXiv preprint arXiv:1808.00177* (2018).
- [10] Pierre-Luc Bacon, Jean Harb, and Doina Precup. 2017. The Option-Critic Architecture. In *Proceedings of the AAAI Conference on Artificial Intelligence*. arXiv:1609.05140.
- [11] Shikhar Bahl, Russell Mendonca, Lili Chen, Unnat Jain, and Deepak Pathak. 2023. Affordances from Human Videos as a Versatile Representation for Robotics. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:2304.08488.
- [12] Homanga Bharadhwaj et al. 2024. RoboAgent: Generalization and Efficiency in Robot Manipulation via Semantic Augmentations and Action Chunking. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2309.01918.
- [13] Kevin Black et al. 2024. Zero-Shot Robotic Manipulation with Pretrained Image-Editing Diffusion Models. In *International Conference on Learning Representations (ICLR)*. arXiv:2310.10639.
- [14] Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Manuel Y. Galliker, Dibya Ghosh, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Devin LeBlanc, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Allen Z. Ren, Lucy Xiaoyang Shi, Laura Smith, Jost Tobias Springenberg, Kyle Stachowicz, James Tanner, Quan Vuong, Homer Walke, Anna Walling, Haohuan Wang, Lili Yu, and Ury Zhilinsky. 2025. $\pi_{0.5}$: A Vision-Language-Action Model with Open-World Generalization. *arXiv preprint* (2025). Physical Intelligence. arXiv:2504.16054.
- [15] Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Lucy Xiaoyang Shi, James Tanner, Quan Vuong, Anna Walling, Haohuan Wang,

- and Ury Zhilinsky. 2024. π_0 : A Vision-Language-Action Flow Model for General Robot Control. *arXiv preprint* (2024). Physical Intelligence. arXiv:2410.24164.
- [16] Konstantinos Bousmalis, Giulia Vezzani, Dushyant Rao, Coline Devin, Alex X. Lee, Maria Bauza, et al. 2023. RoboCat: A Self-Improving Generalist Agent for Robotic Manipulation. *Transactions on Machine Learning Research (TMLR)* (2023). arXiv:2306.11706.
 - [17] Anthony Brohan, Noah Brown, Justice Carbajal, et al. 2023. RT-1: Robotics Transformer for Real-World Control at Scale. In *Robotics: Science and Systems (RSS)*. arXiv:2212.06817.
 - [18] Anthony Brohan, Noah Brown, Justice Carbajal, et al. 2023. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. In *Conference on Robot Learning (CoRL)*. arXiv:2307.15818.
 - [19] Jake Bruce, Michael Dennis, Ashley Edwards, Jack Parker-Holder, et al. 2024. Genie: Generative Interactive Environments. In *International Conference on Machine Learning (ICML)*. arXiv:2402.15391.
 - [20] Victor Campos, Alexander Trott, Caiming Xiong, Richard Socher, Xavier Giro-i Nieto, and Jordi Torres. 2020. Explore, Discover and Learn: Unsupervised Discovery of State-Covering Skills. In *International Conference on Machine Learning (ICML)*. arXiv:2002.03647.
 - [21] Thomas Carta, Clément Romic, Thomas Wolf, Sylvain Lamprier, Olivier Sigaud, and Pierre-Yves Oudeyer. 2023. Grounding Large Language Models in Interactive Environments with Online Reinforcement Learning. *arXiv preprint arXiv:2302.02662* (2023).
 - [22] Chi-Lam Cheang et al. 2024. GR-2: A Generative Video-Language-Action Model with Web-Scale Knowledge for Robot Manipulation. *arXiv preprint arXiv:2410.06158* (2024).
 - [23] Yevgen Chebotar et al. 2023. Q-Transformer: Scalable Offline Reinforcement Learning via Autoregressive Q-Functions. In *Conference on Robot Learning (CoRL)*. arXiv:2309.10150.
 - [24] Yevgen Chebotar, Ankur Handa, Viktor Makoviychuk, Miles Macklin, Jan Issac, Nathan Ratliff, and Dieter Fox. 2019. Closing the Sim-to-Real Loop: Adapting Simulation Randomization with Real World Experience. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:1810.05687.
 - [25] Chang Chen, Yi-Fu Wu, Jaesik Yoon, and Sungjin Ahn. 2022. TransDreamer: Reinforcement Learning with Transformer World Models. *arXiv preprint arXiv:2202.09481* (2022).
 - [26] Harold Haodong Chen, Sirui Chen, Yingjie Xu, Wenhao Ge, and Ying-Cong Chen. 2026. RoboEvolve: Co-Evolving Planner-Simulator for Robotic Manipulation with Limited Data. *arXiv preprint* (2026). arXiv:2605.13775.
 - [27] Juntong Chen, Yao Mu, Qiaojun Yu, Tianming Wei, Silang Wu, Zhecheng Yuan, Zhixuan Liang, Chao Yang, Kaipeng Zhang, Wenqi Shao, Yu Qiao, Huazhe Xu, Mingyu Ding, and Ping Luo. 2024. RoboScript: Code Generation for Free-Form Manipulation Tasks across Real and Simulation. *arXiv preprint* (2024). arXiv:2402.14623.
 - [28] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Michael Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. 2021. Decision Transformer: Reinforcement Learning via Sequence Modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2106.01345.
 - [29] Lawrence Yunliang Chen, Kush Hari, Karthik Dharmarajan, Chenfeng Xu, Quan Vuong, and Ken Goldberg. 2024. Mirage: Cross-Embodiment Zero-Shot Policy Transfer with Cross-Painting. In *Robotics: Science and Systems (RSS)*. arXiv:2402.19249.
 - [30] Tao Chen, Megha Tippur, Siyang Wu, Vikash Kumar, Edward Adelson, and Pulkit Agrawal. 2023. Visual Dexterity: In-Hand Reorientation of Novel and Complex Object Shapes. *Science Robotics* (2023). arXiv:2211.11744.
 - [31] Wentse Chen, Shiyu Huang, Yuan Chiang, Ting Chen, and Jun Zhu. 2024. DGPO: Discovering Multiple Strategies with Diversity-Guided Policy Optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*. arXiv:2207.05631.
 - [32] Yongchao Chen, Jacob Arkin, Yang Zhang, Nicholas Roy, and Chuchu Fan. 2023. AutoTAMP: Autoregressive Task and Motion Planning with LLMs as Translators and Checkers. *arXiv preprint arXiv:2306.06531* (2023).
 - [33] Yongchao Chen, Rujul Gandhi, Yang Zhang, and Chuchu Fan. 2023. NL2TL: Transforming Natural Languages to Temporal Logics using Large Language Models. *arXiv preprint arXiv:2305.07766* (2023).
 - [34] Zoey Chen, Sho Kiani, Abhishek Gupta, and Vikash Kumar. 2023. GenAug: Retargeting behaviors to unseen situations via Generative Augmentation. *arXiv preprint arXiv:2302.06671* (2023).
 - [35] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. 2024. Expressive Whole-Body Control for Humanoid Robots. In *Robotics: Science and Systems (RSS)*. arXiv:2402.16796.
 - [36] Xuxin Cheng, Kexin Shi, Ananya Agarwal, and Deepak Pathak. 2024. Extreme Parkour with Legged Robots. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2309.14341.
 - [37] Cheng Chi et al. 2023. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. In *Robotics: Science and Systems (RSS)*. arXiv:2303.04137.
 - [38] Cheng Chi et al. 2024. Universal Manipulation Interface: In-The-Wild Robot Teaching Without In-The-Wild Robots. In *Robotics: Science and Systems (RSS)*. arXiv:2402.10329.

- [39] Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. 2018. Deep Reinforcement Learning in a Handful of Trials using Probabilistic Dynamics Models. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:1805.12114.
- [40] Sudeep Dasari, Frederik Ebert, Stephen Tian, Suraj Nair, Bernadette Bucher, Karl Schmeckpeper, Siddharth Singh, Sergey Levine, and Chelsea Finn. 2019. RoboNet: Large-Scale Multi-Robot Learning. In *Conference on Robot Learning (CoRL)*. arXiv:1910.11215.
- [41] Jingtao Ding et al. 2025. Understanding World or Predicting Future? A Comprehensive Survey of World Models. *Comput. Surveys* (2025). arXiv:2411.14499.
- [42] Yan Ding, Xiaohan Zhang, Saeid Amiri, Nieqing Cao, Hao Yang, et al. 2022. Robot Task Planning and Situation Handling in Open Worlds. *arXiv preprint arXiv:2210.01287* (2022).
- [43] Yan Ding, Xiaohan Zhang, Chris Paxton, and Shiqi Zhang. 2023. Task and Motion Planning with Large Language Models for Object Rearrangement. *arXiv preprint arXiv:2303.06247* (2023).
- [44] Ria Doshi, Homer Walke, Oier Mees, Sudeep Dasari, and Sergey Levine. 2024. Scaling Cross-Embodied Learning: One Policy for Manipulation, Navigation, Locomotion and Aviation. In *Conference on Robot Learning (CoRL)*. arXiv:2408.11812.
- [45] Danny Driess et al. 2023. PaLM-E: An Embodied Multimodal Language Model. In *International Conference on Machine Learning (ICML)*. arXiv:2303.03378.
- [46] Yilun Du et al. 2023. Learning Universal Policies via Text-Guided Video Generation. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2302.00111.
- [47] Yuqing Du, Olivia Watkins, Zihan Wang, Cédric Colas, Trevor Darrell, Pieter Abbeel, Abhishek Gupta, and Jacob Andreas. 2023. Guiding Pretraining in Reinforcement Learning with Large Language Models. *arXiv preprint arXiv:2302.06692* (2023).
- [48] Jiafei Duan, Wilbert Pumacay, Nishanth Kumar, Yi Ru Wang, Shulin Tian, Wentao Yuan, Ranjay Krishna, Dieter Fox, Ajay Mandlekar, and Yijie Guo. 2024. AHA: A Vision-Language-Model for Detecting and Reasoning Over Failures in Robotic Manipulation. *arXiv preprint* (2024). arXiv:2410.00371.
- [49] Ben Eisner, Harry Zhang, and David Held. 2022. FlowBot3D: Learning 3D Articulation Flow to Manipulate Articulated Objects. In *Robotics: Science and Systems (RSS)*. arXiv:2205.04382.
- [50] Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. 2019. Diversity is All You Need: Learning Skills without a Reward Function. In *International Conference on Learning Representations (ICLR)*. arXiv:1802.06070.
- [51] Hao-Shu Fang, Hongjie Fang, Zhenyu Tang, Jirong Liu, Chenxi Wang, Junbo Wang, Haoyi Zhu, and Cewu Lu. 2023. RH20T: A Comprehensive Robotic Dataset for Learning Diverse Skills in One-Shot. *arXiv preprint arXiv:2307.00595* (2023).
- [52] Hao-Shu Fang, Chenxi Wang, Hongjie Fang, Minghao Gou, Jirong Liu, Hengxu Yan, Wenhai Liu, Yichen Xie, and Cewu Lu. 2023. AnyGrasp: Robust and Efficient Grasp Perception in Spatial and Temporal Domains. *IEEE Transactions on Robotics (T-RO)* (2023). arXiv:2212.08333.
- [53] Hao-Shu Fang, Chenxi Wang, Minghao Gou, and Cewu Lu. 2020. GraspNet-1Billion: A Large-Scale Benchmark for General Object Grasping. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR42600.2020.01146.
- [54] Pete Florence et al. 2021. Implicit Behavioral Cloning. In *Conference on Robot Learning (CoRL)*. arXiv:2109.00137.
- [55] Carlos Florensa, Yan Duan, and Pieter Abbeel. 2017. Stochastic Neural Networks for Hierarchical Reinforcement Learning. In *International Conference on Learning Representations (ICLR)*. arXiv:1704.03012.
- [56] Kevin Frans, Jonathan Ho, Xi Chen, Pieter Abbeel, and John Schulman. 2018. Meta Learning Shared Hierarchies. In *International Conference on Learning Representations (ICLR)*. arXiv:1710.09767.
- [57] Justin Fu, Aviral Kumar, Ofir Nachum, George Tucker, and Sergey Levine. 2020. D4RL: Datasets for Deep Data-Driven Reinforcement Learning. *arXiv preprint arXiv:2004.07219* (2020).
- [58] Letian Fu, Justin Yu, Karim El-Refai, Ethan Kou, Haoru Xue, Huang Huang, Wenli Xiao, Guanzhi Wang, Dantong Niu, Li Fei-Fei, Guanya Shi, Jiajun Wu, Shankar Sastry, Yuke Zhu, Ken Goldberg, and Linxi Fan. 2026. CaP-X: A Framework for Benchmarking and Improving Coding Agents for Robot Manipulation. *arXiv preprint* (2026). arXiv:2603.22435.
- [59] Zipeng Fu, Qingqing Zhao, Qi Wu, Gordon Wetzstein, and Chelsea Finn. 2024. HumanPlus: Humanoid Shadowing and Imitation from Humans. In *Conference on Robot Learning (CoRL)*. arXiv:2406.10454.
- [60] Zipeng Fu, Tony Z. Zhao, and Chelsea Finn. 2024. Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation. *arXiv preprint arXiv:2401.02117* (2024).
- [61] Scott Fujimoto and Shixiang Shane Gu. 2021. A Minimalist Approach to Offline Reinforcement Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2106.06860.
- [62] Scott Fujimoto, David Meger, and Doina Precup. 2019. Off-Policy Deep Reinforcement Learning without Exploration. In *International Conference on Machine Learning (ICML)*. arXiv:1812.02900.

- [63] Samir Yitzhak Gadre, Mitchell Wortsman, Gabriel Ilharco, Ludwig Schmidt, and Shuran Song. 2023. CoWs on Pasture: Baselines and Benchmarks for Language-Driven Zero-Shot Object Navigation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:2203.10421.
- [64] Chuang Gan, Jeremy Schwartz, Seth Alter, Martin Schrimpf, James Traer, Julian De Freitas, Jonas Kubilius, Abhishek Bhandwalder, Nick Haber, Megumi Sano, et al. 2020. ThreeDWorld: A Platform for Interactive Multi-Modal Physical Simulation. *arXiv preprint arXiv:2007.04954* (2020).
- [65] Gemini Robotics Team, Google DeepMind. 2025. Gemini Robotics: Bringing AI into the Physical World. *arXiv preprint* (2025). arXiv:2503.20020.
- [66] Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, Jianlan Luo, You Liang Tan, Lawrence Yunliang Chen, Pannag Sanketi, Quan Vuong, Ted Xiao, Dorsa Sadigh, Chelsea Finn, and Sergey Levine. 2024. Octo: An Open-Source Generalist Robot Policy. In *Robotics: Science and Systems (RSS)*. Octo Model Team. arXiv:2405.12213.
- [67] Ankit Goyal et al. 2023. RVT: Robotic View Transformer for 3D Object Manipulation. In *Conference on Robot Learning (CoRL)*. arXiv:2306.14896.
- [68] Ankit Goyal et al. 2024. RVT-2: Learning Precise Manipulation from Few Demonstrations. In *Robotics: Science and Systems (RSS)*. arXiv:2406.08545.
- [69] Kristen Grauman et al. 2022. Ego4D: Around the World in 3,000 Hours of Egocentric Video. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:2110.07058.
- [70] Karol Gregor, Danilo Jimenez Rezende, and Daan Wierstra. 2016. Variational Intrinsic Control. *arXiv preprint arXiv:1611.07507* (2016).
- [71] Jiayuan Gu, Fanbo Xiang, Xuanlin Li, Zhan Ling, Xiqiang Liu, Tongzhou Mu, Yihe Tang, Stone Tao, Xinyue Wei, Yunchao Yao, Xiaodi Yuan, Pengwei Xie, Zhiao Huang, Rui Chen, and Hao Su. 2023. ManiSkill2: A Unified Benchmark for Generalizable Manipulation Skills. In *International Conference on Learning Representations (ICLR)*. arXiv:2302.04659.
- [72] Ping Guo, Chao Li, Yinglan Feng, and Chaoning Zhang. 2026. Code Evolution for Control: Synthesizing Policies via LLM-Driven Evolutionary Search. *arXiv preprint* (2026). arXiv:2601.06845.
- [73] Yanjiang Guo, Yen-Jen Wang, Lihan Zha, and Jianyu Chen. 2024. DoReMi: Grounding Language Model by Detecting and Recovering from Plan-Execution Misalignment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. arXiv:2307.00329.
- [74] Abhishek Gupta, Vikash Kumar, Corey Lynch, Sergey Levine, and Karol Hausman. 2019. Relay Policy Learning: Solving Long-Horizon Tasks via Imitation and Reinforcement Learning. In *Conference on Robot Learning (CoRL)*. arXiv:1910.11956.
- [75] Irmak Guzey, Yinlong Dai, Ben Evans, Soumith Chintala, and Lerrel Pinto. 2023. See to Touch: Learning Tactile Dexterity through Visual Incentives. *arXiv preprint arXiv:2309.12300* (2023).
- [76] Irmak Guzey, Ben Evans, Soumith Chintala, and Lerrel Pinto. 2023. Dexterity from Touch: Self-Supervised Pre-Training of Tactile Representations with Robotic Play. In *Conference on Robot Learning (CoRL)*. arXiv:2303.12076.
- [77] David Ha and Jürgen Schmidhuber. 2018. World Models. *arXiv preprint arXiv:1803.10122* (2018).
- [78] Huy Ha, Pete Florence, and Shuran Song. 2023. Scaling Up and Distilling Down: Language-Guided Robot Skill Acquisition. *arXiv preprint arXiv:2307.14535* (2023).
- [79] Danijar Hafner, Kuang-Huei Lee, Ian Fischer, and Pieter Abbeel. 2022. Deep Hierarchical Planning from Pixels. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2206.04114.
- [80] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. 2020. Dream to Control: Learning Behaviors by Latent Imagination. In *International Conference on Learning Representations (ICLR)*. arXiv:1912.01603.
- [81] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. 2019. Learning Latent Dynamics for Planning from Pixels. In *International Conference on Machine Learning (ICML)*. arXiv:1811.04551.
- [82] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. 2021. Mastering Atari with Discrete World Models. In *International Conference on Learning Representations (ICLR)*. arXiv:2010.02193.
- [83] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. 2023. Mastering Diverse Domains through World Models. *arXiv preprint arXiv:2301.04104* (2023).
- [84] Ankur Handa et al. 2020. DexPilot: Vision-Based Teleoperation of Dexterous Robotic Hand-Arm System. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:1910.03135.
- [85] Nicklas Hansen, Hao Su, and Xiaolong Wang. 2024. TD-MPC2: Scalable, Robust World Models for Continuous Control. In *International Conference on Learning Representations (ICLR)*. arXiv:2310.16828.
- [86] Nicklas Hansen, Xiaolong Wang, and Hao Su. 2022. Temporal Difference Learning for Model Predictive Control. In *International Conference on Machine Learning (ICML)*. arXiv:2203.04955.
- [87] Steven Hansen, Will Dabney, André Barreto, Tom Van de Wiele, David Warde-Farley, and Volodymyr Mnih. 2020. Fast Task Inference with Variational Intrinsic Successor Features. In *International Conference on Learning Representations*

- (ICLR). arXiv:1906.05030.
- [88] Shuncheng He, Yuhang Jiang, Hongchang Zhang, Jianzhun Shao, and Xiangyang Ji. 2022. Wasserstein Unsupervised Reinforcement Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*. arXiv:2110.07940.
 - [89] Tairan He, Jiawei Gao, Wenli Xiao, Yuanhang Zhang, Zi Wang, Jiashun Wang, Zhengyi Luo, Guanqi He, Nikhil Sobanbabu, Chaoyi Pan, et al. 2025. ASAP: Aligning Simulation and Real-World Physics for Learning Agile Humanoid Whole-Body Skills. In *Robotics: Science and Systems (RSS)*. arXiv:2502.01143.
 - [90] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. 2024. OmniH2O: Universal and Dexterous Human-to-Humanoid Whole-Body Teleoperation and Learning. In *Conference on Robot Learning (CoRL)*. arXiv:2406.08858.
 - [91] Tairan He, Zhengyi Luo, Wenli Xiao, Chong Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. 2024. Learning Human-to-Humanoid Real-Time Whole-Body Teleoperation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. arXiv:2403.04436.
 - [92] Tairan He, Wenli Xiao, Toru Lin, Zhengyi Luo, Zhenjia Xu, Zhenyu Jiang, Jan Kautz, Changliu Liu, Guanya Shi, Xiaolong Wang, Linxi Fan, and Yuke Zhu. 2025. HOVER: Versatile Neural Whole-Body Controller for Humanoid Robots. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2410.21229.
 - [93] Mengkang Hu, Yao Mu, Xinmiao Yu, Mingyu Ding, Shiguang Wu, et al. 2023. Tree-Planner: Efficient Close-loop Task Planning with Large Language Models. *arXiv preprint arXiv:2310.08582* (2023).
 - [94] Yucheng Hu et al. 2024. Video Prediction Policy: A Generalist Robot Policy with Predictive Visual Representations. *arXiv preprint arXiv:2412.14803* (2024).
 - [95] Pu Hua, Minghuan Liu, Annabella Macaluso, Yunfeng Lin, Weinan Zhang, Huazhe Xu, and Lirui Wang. 2024. GenSim2: Scaling Robot Data Generation with Multi-modal and Reasoning LLMs. *arXiv preprint arXiv:2410.03645* (2024).
 - [96] Jiangyong Huang et al. 2024. An Embodied Generalist Agent in 3D World. In *International Conference on Machine Learning (ICML)*. arXiv:2311.12871.
 - [97] Siyuan Huang, Zhengkai Jiang, Hao Dong, Yu Qiao, Peng Gao, and Hongsheng Li. 2023. Instruct2Act: Mapping Multi-modality Instructions to Robotic Actions with Large Language Model. *arXiv preprint* (2023). arXiv:2305.11176.
 - [98] Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. 2022. Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents. *arXiv preprint arXiv:2201.07207* (2022).
 - [99] Wenlong Huang, Chen Wang, Yunzhu Li, Ruohan Zhang, and Li Fei-Fei. 2024. ReKep: Spatio-Temporal Reasoning of Relational Keypoint Constraints for Robotic Manipulation. *arXiv preprint arXiv:2409.01652* (2024).
 - [100] Wenlong Huang, Chen Wang, Ruohan Zhang, Yunzhu Li, Jiajun Wu, and Li Fei-Fei. 2023. VoxPoser: Composable 3D Value Maps for Robotic Manipulation with Language Models. In *Conference on Robot Learning (CoRL)*. arXiv:2307.05973.
 - [101] Wenlong Huang, Fei Xia, Dhruv Shah, Danny Driess, Andy Zeng, et al. 2023. Grounded Decoding: Guiding Text Generation with Grounded Models for Embodied Agents. *arXiv preprint arXiv:2303.00855* (2023).
 - [102] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, Pierre Sermanet, Noah Brown, Tomas Jackson, Linda Luu, Sergey Levine, Karol Hausman, and Brian Ichter. 2022. Inner Monologue: Embodied Reasoning through Planning with Language Models. In *Conference on Robot Learning (CoRL)*. arXiv:2207.05608.
 - [103] Jemin Hwangbo, Joonho Lee, Alexey Dosovitskiy, Dario Bellicoso, Vassilios Tsounis, Vladlen Koltun, and Marco Hutter. 2019. Learning agile and dynamic motor skills for legged robots. *Science Robotics* (2019). arXiv:1901.08652.
 - [104] Stephen James et al. 2019. Sim-to-Real via Sim-to-Sim: Data-efficient Robotic Grasping via Randomized-to-Canonical Adaptation Networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:1812.07252.
 - [105] Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J. Davison. 2020. RL-Bench: The Robot Learning Benchmark and Learning Environment. *IEEE Robotics and Automation Letters (RA-L)* (2020). arXiv:1909.12271.
 - [106] Eric Jang et al. 2021. BC-Z: Zero-Shot Task Generalization with Robotic Imitation Learning. In *Conference on Robot Learning (CoRL)*. arXiv:2202.02005.
 - [107] Michael Janner, Yilun Du, Joshua B. Tenenbaum, and Sergey Levine. 2022. Planning with Diffusion for Flexible Behavior Synthesis. In *International Conference on Machine Learning (ICML)*. arXiv:2205.09991.
 - [108] Michael Janner, Justin Fu, Marvin Zhang, and Sergey Levine. 2019. When to Trust Your Model: Model-Based Policy Optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:1906.08253.
 - [109] Michael Janner, Qiyang Li, and Sergey Levine. 2021. Offline Reinforcement Learning as One Big Sequence Modeling Problem. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2106.02039.
 - [110] Yunfan Jiang et al. 2023. VIMA: General Robot Manipulation with Multimodal Prompts. In *International Conference on Machine Learning (ICML)*. arXiv:2210.03094.
 - [111] Zhenyu Jiang, Yuqi Xie, Kevin Lin, Zhenjia Xu, Weikang Wan, Ajay Mandlekar, Linxi Fan, and Yuke Zhu. 2024. DexMimicGen: Automated Data Generation for Bimanual Dexterous Manipulation via Imitation Learning. *arXiv preprint arXiv:2410.24185* (2024).

- [112] Lukasz Kaiser, Mohammad Babaeizadeh, Piotr Milos, Blazej Osinski, et al. 2020. Model-Based Reinforcement Learning for Atari. In *International Conference on Learning Representations (ICLR)*. arXiv:1903.00374.
- [113] Dmitry Kalashnikov et al. 2018. QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation. In *Conference on Robot Learning (CoRL)*. arXiv:1806.10293.
- [114] Dmitry Kalashnikov et al. 2021. MT-Opt: Continuous Multi-Task Robotic Reinforcement Learning at Scale. *arXiv preprint arXiv:2104.08212* (2021).
- [115] Shyam Sundar Kannan, Vishnunandan L. N. Venkatesh, and Byung-Cheol Min. 2023. SMART-LLM: Smart Multi-Agent Robot Task Planning using Large Language Models. *arXiv preprint arXiv:2309.10062* (2023).
- [116] Siddharth Karamcheti, Suraj Nair, Annie S. Chen, Thomas Kollar, Chelsea Finn, Dorsa Sadigh, and Percy Liang. 2023. Language-Driven Representation Learning for Robotics. In *Robotics: Science and Systems (RSS)*. arXiv:2302.12766.
- [117] Tsung-Wei Ke, Nikolaos Gkanatsios, and Katerina Fragkiadaki. 2024. 3D Diffuser Actor: Policy Diffusion with 3D Scene Representations. *arXiv preprint arXiv:2402.10885* (2024).
- [118] Alexander Khazatsky et al. 2024. DROID: A Large-Scale In-The-Wild Robot Manipulation Dataset. In *Robotics: Science and Systems (RSS)*. arXiv:2403.12945.
- [119] Moo Jin Kim, Chelsea Finn, and Percy Liang. 2025. Fine-Tuning Vision-Language-Action Models: Optimizing Speed and Success. *arXiv preprint arXiv:2502.19645* (2025).
- [120] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. 2024. OpenVLA: An Open-Source Vision-Language-Action Model. In *Conference on Robot Learning (CoRL)*. arXiv:2406.09246.
- [121] Martin Klissarov, Pierluca D'Oro, Shagun Sodhani, Roberta Raileanu, Pierre-Luc Bacon, Pascal Vincent, Amy Zhang, and Mikael Henaff. 2023. Motif: Intrinsic Motivation from Artificial Intelligence Feedback. *arXiv preprint arXiv:2310.00166* (2023).
- [122] Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Daniel Gordon, Yuke Zhu, Abhinav Gupta, and Ali Farhadi. 2017. AI2-THOR: An Interactive 3D Environment for Visual AI. *arXiv preprint arXiv:1712.05474* (2017).
- [123] Ilya Kostrikov, Ashvin Nair, and Sergey Levine. 2022. Offline Reinforcement Learning with Implicit Q-Learning. In *International Conference on Learning Representations (ICLR)*. arXiv:2110.06169.
- [124] Tejas D. Kulkarni, Karthik R. Narasimhan, Ardavan Saeedi, and Joshua B. Tenenbaum. 2016. Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:1604.06057.
- [125] Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. 2021. RMA: Rapid Motor Adaptation for Legged Robots. In *Robotics: Science and Systems (RSS)*. arXiv:2107.04034.
- [126] Aviral Kumar, Aurick Zhou, George Tucker, and Sergey Levine. 2020. Conservative Q-Learning for Offline Reinforcement Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2006.04779.
- [127] Teyun Kwon, Norman Di Palo, and Edward Johns. 2023. Language Models as Zero-Shot Trajectory Generators. *arXiv preprint arXiv:2310.11604* (2023).
- [128] Mike Lambeta, Po-Wei Chou, Stephen Tian, Brian Yang, et al. 2020. DIGIT: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor with Application to In-Hand Manipulation. *IEEE Robotics and Automation Letters (RA-L)* (2020). arXiv:2005.14679.
- [129] Michael Laskin, Hao Liu, Xue Bin Peng, Denis Yarats, Aravind Rajeswaran, and Pieter Abbeel. 2022. CIC: Contrastive Intrinsic Control for Unsupervised Skill Discovery. *arXiv preprint* (2022). arXiv:2202.00161.
- [130] Alex X. Lee, Anusha Nagabandi, Pieter Abbeel, and Sergey Levine. 2020. Stochastic Latent Actor-Critic: Deep Reinforcement Learning with a Latent Variable Model. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:1907.00953.
- [131] Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. 2020. Learning Quadrupedal Locomotion over Challenging Terrain. *Science Robotics* (2020). arXiv:2010.11251.
- [132] Andrew Levy, George Konidaris, Robert Platt, and Kate Saenko. 2019. Learning Multi-Level Hierarchies with Hindsight. In *International Conference on Learning Representations (ICLR)*. arXiv:1712.00948.
- [133] Boyi Li, Philipp Wu, Pieter Abbeel, and Jitendra Malik. 2023. Interactive Task Planning with Language Models. *arXiv preprint arXiv:2310.10645* (2023).
- [134] Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, et al. 2024. BEHAVIOR-1K: A Human-Centered, Embodied AI Benchmark with 1,000 Everyday Activities and Realistic Simulation. *arXiv preprint* (2024). arXiv:2403.09227.
- [135] Dingzhe Li et al. 2024. What Foundation Models can Bring for Robot Learning in Manipulation: A Survey. *arXiv preprint* (2024). arXiv:2404.18201.

- [136] Hao Li, Xue Yang, Zhaokai Wang, Xizhou Zhu, Jie Zhou, Yu Qiao, Xiaogang Wang, Hongsheng Li, Lewei Lu, and Jifeng Dai. 2024. Auto MC-Reward: Automated Dense Reward Design with Large Language Models for Minecraft. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:2312.09238.
- [137] Jingyao Li, Pengguang Chen, Sitong Wu, Chuanyang Zheng, Hong Xu, and Jiaya Jia. 2024. RoboCoder: Robotic Learning from Basic Skills to General Tasks with Large Language Models. *arXiv preprint* (2024). arXiv:2406.03757.
- [138] Qixiu Li, Yaobo Liang, Zeyu Wang, Lin Luo, Xi Chen, Mozheng Liao, Fangyun Wei, Yu Deng, Sicheng Xu, Yizhong Zhang, Xiaofan Wang, Bei Liu, Jianlong Fu, Jianmin Bao, Dong Chen, Yuanchun Shi, Jiaolong Yang, and Baining Guo. 2024. CogACT: A Foundational Vision-Language-Action Model for Synergizing Cognition and Action in Robotic Manipulation. *arXiv preprint* (2024). arXiv:2411.19650.
- [139] Ruiying Li, Yunlang Zhou, YuYao Zhu, Kylin Chen, Jingyuan Wang, Sukai Wang, Kongtao Hu, Minhui Yu, Bowen Jiang, Zhan Su, Jiayao Ma, Xin He, Yongjian Shen, Yang Yang, Guanghui Ren, Maoqing Yao, Wenhao Wang, and Yao Mu. 2026. RoboClaw: An Agentic Framework for Scalable Long-Horizon Robotic Tasks. *arXiv preprint* (2026). arXiv:2603.11558.
- [140] Xiang Li et al. 2024. LLaRA: Supercharging Robot Learning Data for Vision-Language Policy. *arXiv preprint arXiv:2406.20095* (2024).
- [141] Xiaoqi Li et al. 2024. ManipLLM: Embodied Multimodal Large Language Model for Object-Centric Robotic Manipulation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:2312.16217.
- [142] Xinghang Li, Minghuan Liu, Hanbo Zhang, Cunjun Yu, Jie Xu, Hongtao Wu, Chilam Cheang, Ya Jing, Weinan Zhang, Huaping Liu, Hang Li, and Tao Kong. 2024. Vision-Language Foundation Models as Effective Robot Imitators. In *International Conference on Learning Representations (ICLR)*. RoboFlamingo. arXiv:2311.01378.
- [143] Zaijing Li, Yuquan Xie, Rui Shao, Gongwei Chen, Dongmei Jiang, and Liqiang Nie. 2024. Optimus-1: Hybrid Multimodal Memory Empowered Agents Excel in Long-Horizon Tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2408.03615.
- [144] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. 2023. Code as Policies: Language Model Programs for Embodied Control. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2209.07753.
- [145] Kaiqi Liang, Zixu Zhang, and Jaime Fernández Fisac. 2024. Introspective Planning: Aligning Robots' Uncertainty with Inherent Task Ambiguity. *arXiv preprint* (2024). arXiv:2402.06529.
- [146] William Liang, Sam Wang, Hung-Ju Wang, Osbert Bastani, Dinesh Jayaraman, and Yecheng Jason Ma. 2024. Eureka-verse: Environment Curriculum Generation via Large Language Models. In *Conference on Robot Learning (CoRL)*. arXiv:2411.01775.
- [147] Kevin Lin, Christopher Agia, Toki Migimatsu, Marco Pavone, and Jeannette Bohg. 2023. Text2Motion: From Natural Language Instructions to Feasible Plans. *Autonomous Robots* (2023). arXiv:2303.12153.
- [148] Bo Liu, Yuqian Jiang, Xiaohan Zhang, Qiang Liu, Shiqi Zhang, Joydeep Biswas, and Peter Stone. 2023. LLM+P: Empowering Large Language Models with Optimal Planning Proficiency. *arXiv preprint arXiv:2304.11477* (2023).
- [149] Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. 2023. LIBERO: Benchmarking Knowledge Transfer for Lifelong Robot Learning. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*. arXiv:2306.03310.
- [150] Fangchen Liu, Kuan Fang, Pieter Abbeel, and Sergey Levine. 2024. MOKA: Open-World Robotic Manipulation through Mark-Based Visual Prompting. *arXiv preprint arXiv:2403.03174* (2024).
- [151] Hao Liu and Pieter Abbeel. 2021. APS: Active Pretraining with Successor Features. In *International Conference on Machine Learning (ICML)*. arXiv:2108.13956.
- [152] Hao Liu and Pieter Abbeel. 2021. Behavior From the Void: Unsupervised Active Pre-Training. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2103.04551.
- [153] Huaping Liu, Di Guo, and Angelo Cangelosi. 2025. Embodied Intelligence: A Synergy of Morphology, Action, Perception and Learning. *Comput. Surveys* (2025). DOI:10.1145/3717059.
- [154] Jiaming Liu et al. 2024. RoboMamba: Multimodal State Space Model for Efficient Robot Reasoning and Manipulation. *arXiv preprint arXiv:2406.04339* (2024).
- [155] Jason Xinyu Liu, Ziyi Yang, Ifrah Idrees, Sam Liang, Benjamin Schornstein, Stefanie Tellex, and Ankit Shah. 2023. Grounding Complex Natural Language Commands for Temporal Tasks in Unseen Environments. *arXiv preprint arXiv:2302.11649* (2023).
- [156] Songming Liu et al. 2024. RDT-1B: a Diffusion Foundation Model for Bimanual Manipulation. *arXiv preprint arXiv:2410.07864* (2024).
- [157] Shunyu Liu, Yaoru Li, Kongcheng Zhang, Zhenyu Cui, Wenkai Fang, Yuxuan Zheng, Tongya Zheng, and Mingli Song. 2024. Odyssey: Empowering Minecraft Agents with Open-World Skills. *arXiv preprint* (2024). arXiv:2407.15325.
- [158] Yuchen Liu, Luigi Palmieri, Sebastian Koch, Ilche Georgievski, and Marco Aiello. 2024. DELTA: Decomposed Efficient Long-Term Robot Task Planning using Large Language Models. *arXiv preprint arXiv:2404.03275* (2024).

- [159] Zeyi Liu, Arpit Bahety, and Shuran Song. 2023. REFLECT: Summarizing Robot Experiences for Failure Explanation and Correction. In *Conference on Robot Learning (CoRL)*. arXiv:2306.15724.
- [160] Runyu Lu, Yubo Wu, Ethan Kou, Letian Fu, Wenli Xiao, Ajay Mandlekar, Yinzen Xu, Guanya Shi, Ken Goldberg, Ang Chen, Mosharaf Chowdhury, Yuke Zhu, Linxi Fan, and Guanzhi Wang. 2026. ASPIRE: Agentic Skills Discovery for Robotics. *arXiv preprint* (2026). arXiv:2607.00272.
- [161] Yueen Ma et al. 2024. A Survey on Vision-Language-Action Models for Embodied AI. *arXiv preprint* (2024). arXiv:2405.14093.
- [162] Yecheng Jason Ma, William Liang, Vaidehi Som, Vikash Kumar, Amy Zhang, Osbert Bastani, and Dinesh Jayaraman. 2023. LIV: Language-Image Representations and Rewards for Robotic Control. In *International Conference on Machine Learning (ICML)*. arXiv:2306.00958.
- [163] Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2024. Eureka: Human-Level Reward Design via Coding Large Language Models. In *International Conference on Learning Representations (ICLR)*. arXiv:2310.12931.
- [164] Yecheng Jason Ma, William Liang, Hung-Ju Wang, Sam Wang, Yuke Zhu, Linxi Fan, Osbert Bastani, and Dinesh Jayaraman. 2024. DrEureka: Language Model Guided Sim-To-Real Transfer. In *Robotics: Science and Systems (RSS)*. arXiv:2406.01967.
- [165] Yecheng Jason Ma, Shagun Sodhani, Dinesh Jayaraman, Osbert Bastani, Vikash Kumar, and Amy Zhang. 2023. VIP: Towards Universal Visual Reward and Representation via Value-Implicit Pre-Training. In *International Conference on Learning Representations (ICLR)*. arXiv:2210.00030.
- [166] Marlos C. Machado, Clemens Rosenbaum, Xiaoxiao Guo, Miao Liu, Gerald Tesauro, and Murray Campbell. 2018. Eigenoption Discovery through the Deep Successor Representation. In *International Conference on Learning Representations (ICLR)*. arXiv:1710.11089.
- [167] Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, et al. 2023. Self-Refine: Iterative Refinement with Self-Feedback. *arXiv preprint arXiv:2303.17651* (2023).
- [168] Jeffrey Mahler et al. 2017. Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics. In *Robotics: Science and Systems (RSS)*. arXiv:1703.09312.
- [169] Arjun Majumdar, Karmesh Yadav, Sergio Arnaud, Yecheng Jason Ma, et al. 2023. Where are we in the search for an Artificial Visual Cortex for Embodied Intelligence?. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2303.18240.
- [170] Viktor Makovychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. 2021. Isaac Gym: High Performance GPU-Based Physics Simulation For Robot Learning. *arXiv preprint arXiv:2108.10470* (2021).
- [171] Zhao Mandi, Homanga Bharadhwaj, Vincent Moens, Shuran Song, Aravind Rajeswaran, and Vikash Kumar. 2022. CACTI: A Framework for Scalable Multi-Task Multi-Scene Visual Imitation Learning. *arXiv preprint arXiv:2212.05711* (2022).
- [172] Zhao Mandi, Shreya Jain, and Shuran Song. 2023. RoCo: Dialectic Multi-Robot Collaboration with Large Language Models. *arXiv preprint arXiv:2307.04738* (2023).
- [173] Ajay Mandlekar et al. 2021. What Matters in Learning from Offline Human Demonstrations for Robot Manipulation. In *Conference on Robot Learning (CoRL)*. arXiv:2108.03298.
- [174] Ajay Mandlekar et al. 2023. MimicGen: A Data Generation System for Scalable Robot Learning using Human Demonstrations. In *Conference on Robot Learning (CoRL)*. arXiv:2310.17596.
- [175] Gabriel B. Margolis and Pulkit Agrawal. 2022. Walk These Ways: Tuning Robot Control for Generalization with Multiplicity of Behavior. In *Conference on Robot Learning (CoRL)*. arXiv:2212.03238.
- [176] Pietro Mazzaglia, Tim Verbelen, Bart Dhoedt, Alexandre Lacoste, and Sai Rajeswar. 2023. Choreographer: Learning and Adapting Skills in Imagination. In *International Conference on Learning Representations (ICLR)*. arXiv:2211.13350.
- [177] Oier Mees, Lukas Hermann, and Wolfram Burgard. 2022. What Matters in Language Conditioned Robotic Imitation Learning over Unstructured Data. *IEEE Robotics and Automation Letters (RA-L)* (2022). arXiv:2204.06252.
- [178] Oier Mees, Lukas Hermann, Erick Rosete-Beas, and Wolfram Burgard. 2022. CALVIN: A Benchmark for Language-Conditioned Policy Learning for Long-Horizon Robot Manipulation Tasks. *IEEE Robotics and Automation Letters (RA-L)* (2022). arXiv:2112.03227.
- [179] Russell Mendonca, Shikhar Bahl, and Deepak Pathak. 2023. Structured World Models from Human Videos. In *Robotics: Science and Systems (RSS)*. arXiv:2308.10901.
- [180] Josh Merel, Leonard Hasenclever, Alexandre Galashov, Arun Ahuja, Vu Pham, Greg Wayne, Yee Whye Teh, and Nicolas Heess. 2019. Neural Probabilistic Motor Primitives for Humanoid Control. In *International Conference on Learning Representations (ICLR)*. arXiv:1811.11711.
- [181] Vincent Micheli, Eloi Alonso, and François Fleuret. 2023. Transformers are Sample-Efficient World Models. In *International Conference on Learning Representations (ICLR)*. arXiv:2209.00588.

- [182] Takahiro Miki, Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. 2022. Learning robust perceptive locomotion for quadrupedal robots in the wild. *Science Robotics* (2022). arXiv:2201.08117.
- [183] Mayank Mittal, Calvin Yu, Qinxin Yu, Jingzhou Liu, Nikita Rudin, David Hoeller, Jia Lin Yuan, Ritvik Singh, Yunrong Guo, Hammad Mazhar, Ajay Mandlekar, Buck Babich, Gavriel State, Marco Hutter, and Animesh Garg. 2023. Orbit: A Unified Simulation Framework for Interactive Robot Learning Environments. *IEEE Robotics and Automation Letters (RA-L)* (2023). arXiv:2301.04195.
- [184] Kaichun Mo, Leonidas J. Guibas, Mustafa Mukadam, Abhinav Gupta, and Shubham Tulsiani. 2021. Where2Act: From Pixels to Actions for Articulated 3D Objects. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. arXiv:2101.02692.
- [185] Tongzhou Mu, Zhan Ling, Fanbo Xiang, Derek Yang, Xuanlin Li, Stone Tao, Zhiao Huang, Zhiwei Jia, and Hao Su. 2021. ManiSkill: Generalizable Manipulation Skill Benchmark with Large-Scale Demonstrations. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks*. arXiv:2107.14483.
- [186] Yao Mu, Junting Chen, Qinglong Zhang, Shoufa Chen, Qiaojun Yu, Chongjian Ge, Runjian Chen, Zhixuan Liang, Mengkang Hu, Chaofan Tao, Peize Sun, Haibao Yu, Chao Yang, Wenqi Shao, Wenhai Wang, Jifeng Dai, Yu Qiao, Mingyu Ding, and Ping Luo. 2024. RoboCodeX: Multimodal Code Generation for Robotic Behavior Synthesis. In *International Conference on Machine Learning (ICML)*. arXiv:2402.16117.
- [187] Ofir Nachum, Shixiang Gu, Honglak Lee, and Sergey Levine. 2018. Data-Efficient Hierarchical Reinforcement Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:1805.08296.
- [188] Ashvin Nair, Abhishek Gupta, Murtaza Dalal, and Sergey Levine. 2020. AWAC: Accelerating Online Reinforcement Learning with Offline Datasets. *arXiv preprint arXiv:2006.09359* (2020).
- [189] Suraj Nair, Aravind Rajeswaran, Vikash Kumar, Chelsea Finn, and Abhinav Gupta. 2022. R3M: A Universal Visual Representation for Robot Manipulation. In *Conference on Robot Learning (CoRL)*. arXiv:2203.12601.
- [190] Mitsuhiro Nakamoto, Yuexiang Zhai, Anikait Singh, Max Sobol Mark, Yi Ma, Chelsea Finn, Aviral Kumar, and Sergey Levine. 2023. Cal-QL: Calibrated Offline RL Pre-Training for Efficient Online Fine-Tuning. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2303.05479.
- [191] Soroush Nasiriany, Abhiram Maddukuri, Lance Zhang, Adeet Parikh, Aaron Lo, Abhishek Joshi, Ajay Mandlekar, and Yuke Zhu. 2024. RoboCasa: Large-Scale Simulation of Everyday Tasks for Generalist Robots. In *Robotics: Science and Systems (RSS)*. arXiv:2406.02523.
- [192] Soroush Nasiriany, Fei Xia, Wenhao Yu, Ted Xiao, Jacky Liang, et al. 2024. PIVOT: Iterative Visual Prompting Elicits Actionable Knowledge for VLMs. *arXiv preprint arXiv:2402.07872* (2024).
- [193] NVIDIA. 2025. Cosmos World Foundation Model Platform for Physical AI. *arXiv preprint arXiv:2501.03575* (2025).
- [194] NVIDIA, Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Fan, Yu Fang, Dieter Fox, Fengyuan Hu, et al. 2025. GR00T N1: An Open Foundation Model for Generalist Humanoid Robots. *arXiv preprint* (2025). NVIDIA. arXiv:2503.14734.
- [195] Open X-Embodiment Collaboration. 2024. Open X-Embodiment: Robotic Learning Datasets and RT-X Models. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2310.08864.
- [196] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, et al. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (2021), n71. doi:10.1136/bmj.n71.
- [197] Haotian Pan et al. 2026. Robot Navigation via Foundation Language Models: A Review. *Comput. Surveys* (2026). DOI:10.1145/3802539.
- [198] Minting Pan, Xiangming Zhu, Yunbo Wang, and Xiaokang Yang. 2022. Iso-Dream: Isolating and Leveraging Non-controllable Visual Dynamics in World Models. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2205.13817.
- [199] Simone Parisi, Aravind Rajeswaran, Senthil Purushwalkam, and Abhinav Gupta. 2022. The Unsurprising Effectiveness of Pre-Trained Vision Models for Control. In *International Conference on Machine Learning (ICML)*. arXiv:2203.03580.
- [200] Seohong Park, Jongwook Choi, Jaekyeom Kim, Honglak Lee, and Gunhee Kim. 2022. Lipschitz-constrained Unsupervised Skill Discovery. In *International Conference on Learning Representations (ICLR)*. arXiv:2202.00914.
- [201] Seohong Park, Kimin Lee, Youngwoon Lee, and Pieter Abbeel. 2023. Controllability-Aware Unsupervised Skill Discovery. In *International Conference on Machine Learning (ICML)*. arXiv:2302.05103.
- [202] Seohong Park, Oleh Rybkin, and Sergey Levine. 2024. METRA: Scalable Unsupervised RL with Metric-Aware Abstraction. In *International Conference on Learning Representations (ICLR)*. arXiv:2310.08887.
- [203] Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. 2018. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:1710.06537.
- [204] Xue Bin Peng, Michael Chang, Grace Zhang, Pieter Abbeel, and Sergey Levine. 2019. MCP: Learning Composable Hierarchical Control with Multiplicative Compositional Policies. In *Advances in Neural Information Processing Systems*

- (*NeurIPS*). arXiv:1905.09808.
- [205] Karl Pertsch et al. 2025. FAST: Efficient Action Tokenization for Vision-Language-Action Models. *arXiv preprint arXiv:2501.09747* (2025).
 - [206] Karl Pertsch, Youngwoon Lee, and Joseph J. Lim. 2020. Accelerating Reinforcement Learning with Learned Skill Priors. In *Conference on Robot Learning (CoRL)*. arXiv:2010.11944.
 - [207] Delin Qu, Haoming Song, Qizhi Chen, Yuanqi Yao, Xinyi Ye, Yan Ding, Zhigang Wang, JiaYuan Gu, Bin Zhao, Dong Wang, and Xuelong Li. 2025. SpatialVLA: Exploring Spatial Representations for Visual-Language-Action Model. *arXiv preprint* (2025). arXiv:2501.15830.
 - [208] Ilija Radosavovic, Tete Xiao, Bike Zhang, Trevor Darrell, Jitendra Malik, and Koushil Sreenath. 2024. Real-World Humanoid Locomotion with Reinforcement Learning. *Science Robotics* (2024). arXiv:2303.03381.
 - [209] Aravind Rajeswaran, Vikash Kumar, Abhishek Gupta, Giulia Vezzani, John Schulman, Emanuel Todorov, and Sergey Levine. 2018. Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations. In *Robotics: Science and Systems (RSS)*. arXiv:1709.10087.
 - [210] Krishan Rana, Jesse Haviland, Sourav Garg, Jad Abou-Chakra, Ian Reid, and Niko Sünderhauf. 2023. SayPlan: Grounding Large Language Models using 3D Scene Graphs for Scalable Robot Task Planning. *arXiv preprint arXiv:2307.06135* (2023).
 - [211] Scott Reed et al. 2022. A Generalist Agent. *arXiv preprint arXiv:2205.06175* (2022).
 - [212] Allen Z. Ren, Anushri Dixit, Alexandra Bodrova, Sumeet Singh, Stephen Tu, et al. 2023. Robots That Ask For Help: Uncertainty Alignment for Large Language Model Planners. *arXiv preprint arXiv:2307.01928* (2023).
 - [213] Moritz Reuss et al. 2024. Multimodal Diffusion Transformer: Learning Versatile Behavior from Multimodal Goals. In *Robotics: Science and Systems (RSS)*. arXiv:2407.05996.
 - [214] Moritz Reuss, Maximilian Li, Xiaogang Jia, and Rudolf Lioutikov. 2023. Goal-Conditioned Imitation Learning using Score-based Diffusion Policies. In *Robotics: Science and Systems (RSS)*. arXiv:2304.02532.
 - [215] Jan Robine, Marc Höftmann, Tobias Uelwer, and Stefan Harmeling. 2023. Transformer-based World Models Are Happy With 100k Interactions. In *International Conference on Learning Representations (ICLR)*. arXiv:2303.07109.
 - [216] Juan Rocamonde, Victoriano Montesinos, Elvis Nava, Ethan Perez, and David Lindner. 2023. Vision-Language Models are Zero-Shot Reward Models for Reinforcement Learning. *arXiv preprint arXiv:2310.12921* (2023).
 - [217] Nikita Rudin, David Hoeller, Philipp Reist, and Marco Hutter. 2021. Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning. In *Conference on Robot Learning (CoRL)*. arXiv:2109.11978.
 - [218] Fereshteh Sadeghi and Sergey Levine. 2017. CAD2RL: Real Single-Image Flight without a Single Real Image. In *Robotics: Science and Systems (RSS)*. arXiv:1611.04201.
 - [219] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. 2019. Habitat: A Platform for Embodied AI Research. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. arXiv:1904.01201.
 - [220] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, et al. 2020. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. *Nature* (2020). arXiv:1911.08265.
 - [221] Younggyo Seo, Danijar Hafner, Hao Liu, Fangchen Liu, Stephen James, Kimin Lee, and Pieter Abbeel. 2022. Masked World Models for Visual Control. In *Conference on Robot Learning (CoRL)*. arXiv:2206.14244.
 - [222] Nur Muhammad Mahi Shafiullah, Anant Rai, Haritheja Etukuru, Yiqian Liu, Ishan Misra, Soumith Chintala, and Lerrel Pinto. 2023. On Bringing Robots Home. *arXiv preprint arXiv:2311.16098* (2023).
 - [223] Dhruv Shah and Sergey Levine. 2022. ViKiNG: Vision-Based Kilometer-Scale Navigation with Geographic Hints. In *Robotics: Science and Systems (RSS)*. arXiv:2202.11271.
 - [224] Dhruv Shah, Blazej Osinski, Brian Ichter, and Sergey Levine. 2022. LM-Nav: Robotic Navigation with Large Pre-Trained Models of Language, Vision, and Action. In *Conference on Robot Learning (CoRL)*. arXiv:2207.04429.
 - [225] Dhruv Shah, Ajay Sridhar, Arjun Bhorkar, Noriaki Hirose, and Sergey Levine. 2023. GNM: A General Navigation Model to Drive Any Robot. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2210.03370.
 - [226] Dhruv Shah, Ajay Sridhar, Nitish Dashora, Kyle Stachowicz, Kevin Black, Noriaki Hirose, and Sergey Levine. 2023. ViNT: A Foundation Model for Visual Navigation. In *Conference on Robot Learning (CoRL)*. arXiv:2306.14846.
 - [227] Jinghuan Shang, Karl Schmeckpeper, Brandon B. May, Maria Vittoria Minniti, Tarik Kelestemur, David Watkins, and Laura Herlant. 2024. Theia: Distilling Diverse Vision Foundation Models for Robot Learning. In *Conference on Robot Learning (CoRL)*. arXiv:2407.20179.
 - [228] Rui Shao et al. 2025. Large VLM-based Vision-Language-Action Models for Robotic Manipulation: A Survey. *arXiv preprint* (2025). arXiv:2508.13073.
 - [229] Archit Sharma, Shixiang Gu, Sergey Levine, Vikash Kumar, and Karol Hausman. 2020. Dynamics-Aware Unsupervised Discovery of Skills. In *International Conference on Learning Representations (ICLR)*. arXiv:1907.01657.
 - [230] Bokui Shen, Fei Xia, Chengshu Li, Roberto Martín-Martín, Linxi Fan, Guanzhi Wang, Claudia Pérez-D’Arpino, Shyamal Buch, Sanjana Srivastava, Lyne Tchampti, et al. 2021. iGibson 1.0: A Simulation Environment for Interactive Tasks in

- Large Realistic Scenes. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. arXiv:2012.02924.
- [231] Lucy Xiaoyang Shi, Joseph J. Lim, and Youngwoon Lee. 2022. Skill-based Model-based Reinforcement Learning. In *Conference on Robot Learning (CoRL)*. arXiv:2207.07560.
 - [232] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language Agents with Verbal Reinforcement Learning. *arXiv preprint arXiv:2303.11366* (2023).
 - [233] Mohit Shridhar, Lucas Manuelli, and Dieter Fox. 2021. CLIPort: What and Where Pathways for Robotic Manipulation. In *Conference on Robot Learning (CoRL)*. arXiv:2109.12098.
 - [234] Mohit Shridhar, Lucas Manuelli, and Dieter Fox. 2022. Perceiver-Actor: A Multi-Task Transformer for Robotic Manipulation. In *Conference on Robot Learning (CoRL)*. arXiv:2209.05451.
 - [235] Avi Singh, Huihan Liu, Gaoyue Zhou, Albert Yu, Nicholas Rhinehart, and Sergey Levine. 2021. Parrot: Data-Driven Behavioral Priors for Reinforcement Learning. In *International Conference on Learning Representations (ICLR)*. arXiv:2011.10024.
 - [236] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. 2023. ProgPrompt: Generating Situated Robot Task Plans using Large Language Models. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2209.11302.
 - [237] Chan Hee Song, Jiaman Wu, Clayton Washington, Brian M. Sadler, Wei-Lun Chao, and Yu Su. 2022. LLM-Planner: Few-Shot Grounded Planning for Embodied Agents with Large Language Models. *arXiv preprint arXiv:2212.04088* (2022).
 - [238] Ajay Sridhar, Dhruv Shah, Catherine Glossop, and Sergey Levine. 2024. NoMaD: Goal Masked Diffusion Policies for Navigation and Exploration. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2310.07896.
 - [239] Martin Sundermeyer, Arsalan Mousavian, Rudolph Triebel, and Dieter Fox. 2021. Contact-GraspNet: Efficient 6-DoF Grasp Generation in Cluttered Scenes. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2103.14127.
 - [240] Yang Tian et al. 2025. Predictive Inverse Dynamics Models are Scalable Learners for Robotic Manipulation. In *International Conference on Learning Representations (ICLR)*. arXiv:2412.15109.
 - [241] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. 2017. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. arXiv:1703.06907.
 - [242] Georgios Tzafas and Hamidreza Kasaei. 2024. Lifelong Robot Library Learning: Bootstrapping Composable and Generalizable Skills for Embodied Control with Language Models. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2406.18746.
 - [243] Unitree Robotics. 2026. UniStore: A Humanoid Robot Application Store. <https://unistore.unitree.com>.
 - [244] Sai Vemprala, Rogerio Bonatti, Arthur Buckner, and Ashish Kapoor. 2023. ChatGPT for Robotics: Design Principles and Model Abilities. *arXiv preprint* (2023). arXiv:2306.17582.
 - [245] David Venuto, Sami Nur Islam, Martin Klissarov, Doina Precup, Sherry Yang, and Ankit Anand. 2024. Code as Reward: Empowering Reinforcement Learning with VLMs. *arXiv preprint arXiv:2402.04764* (2024).
 - [246] Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. 2017. FeUdal Networks for Hierarchical Reinforcement Learning. In *International Conference on Machine Learning (ICML)*. arXiv:1703.01161.
 - [247] Homer Walke et al. 2023. BridgeData V2: A Dataset for Robot Learning at Scale. In *Conference on Robot Learning (CoRL)*. arXiv:2308.12952.
 - [248] Weikang Wan, Yifeng Zhu, Rutav Shah, and Yuke Zhu. 2024. LOTUS: Continual Imitation Learning for Robot Manipulation Through Unsupervised Skill Discovery. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2311.02058.
 - [249] Dian Wang et al. 2024. Equivariant Diffusion Policy. In *Conference on Robot Learning (CoRL)*. arXiv:2407.01812.
 - [250] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023. Voyager: An Open-Ended Embodied Agent with Large Language Models. *arXiv preprint* (2023). arXiv:2305.16291.
 - [251] Huaxiaoyue Wang, Gonzalo Gonzalez-Pumariaga, Yash Sharma, and Sanjiban Choudhury. 2023. Demo2Code: From Summarizing Demonstrations to Synthesizing Code via Extended Chain-of-Thought. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2305.16744.
 - [252] Lirui Wang, Xinlei Chen, Jialiang Zhao, and Kaiming He. 2024. Scaling Proprioceptive-Visual Learning with Heterogeneous Pre-trained Transformers. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2409.20537.
 - [253] Lirui Wang, Yiyang Ling, Zhecheng Yuan, Mohit Shridhar, Chen Bao, Yuzhe Qin, Bailin Wang, Huazhe Xu, and Xiaolong Wang. 2023. GenSim: Generating Robotic Simulation Tasks via Large Language Models. *arXiv preprint arXiv:2310.01361* (2023).
 - [254] Shu Wang, Muzhi Han, Ziyuan Jiao, Zeyu Zhang, Ying Nian Wu, Song-Chun Zhu, and Hangxin Liu. 2024. LLM3: Large Language Model-based Task and Motion Planning with Motion Failure Reasoning. *arXiv preprint arXiv:2403.11552*

- (2024).
- [255] Yufei Wang, Zhanyi Sun, Jesse Zhang, Zhou Xian, Erdem Biyik, David Held, and Zackory Erickson. 2024. RL-VLM-F: Reinforcement Learning from Vision Language Foundation Model Feedback. *arXiv preprint arXiv:2402.03681* (2024).
 - [256] Yufei Wang, Zhou Xian, Feng Chen, Tsun-Hsuan Wang, Yian Wang, Katerina Fragkiadaki, Zackory Erickson, David Held, and Chuang Gan. 2024. RoboGen: Towards Unleashing Infinite Data for Automated Robot Learning via Generative Simulation. In *International Conference on Machine Learning (ICML)*. arXiv:2311.01455.
 - [257] Yen-Jen Wang, Bike Zhang, Jianyu Chen, and Koushil Sreenath. 2024. Prompt a Robot to Walk with Large Language Models. In *IEEE Conference on Decision and Control (CDC)*. arXiv:2309.09969.
 - [258] Zihao Wang, Shaofei Cai, Guanzhou Chen, Anji Liu, Xiaojian Ma, and Yitao Liang. 2023. Describe, Explain, Plan and Select: Interactive Planning with Large Language Models Enables Open-World Multi-Task Agents. *arXiv preprint arXiv:2302.01560* (2023).
 - [259] Zihao Wang, Shaofei Cai, Anji Liu, Yonggang Jin, Jinbing Hou, Bowei Zhang, Haowei Lin, Zhaofeng He, Zilong Zheng, Yaodong Yang, Xiaojian Ma, and Yitao Liang. 2023. JARVIS-1: Open-World Multi-task Agents with Memory-Augmented Multimodal Language Models. *arXiv preprint* (2023). arXiv:2311.05997.
 - [260] Zhendong Wang, Jonathan J. Hunt, and Mingyuan Zhou. 2023. Diffusion Policies as an Expressive Policy Class for Offline Reinforcement Learning. In *International Conference on Learning Representations (ICLR)*. arXiv:2208.06193.
 - [261] Chuan Wen et al. 2024. Any-point Trajectory Modeling for Policy Learning. In *Robotics: Science and Systems (RSS)*. arXiv:2401.00025.
 - [262] Junjie Wen et al. 2024. TinyVLA: Towards Fast, Data-Efficient Vision-Language-Action Models for Robotic Manipulation. *arXiv preprint arXiv:2409.12514* (2024).
 - [263] Junjie Wen et al. 2025. DexVLA: Vision-Language Model with Plug-In Diffusion Expert for General Robot Control. *arXiv preprint arXiv:2502.05855* (2025).
 - [264] Hongtao Wu et al. 2024. Unleashing Large-Scale Video Generative Pre-training for Visual Robot Manipulation. In *International Conference on Learning Representations (ICLR)*. arXiv:2312.13139.
 - [265] Jimmy Wu, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon Rusinkiewicz, and Thomas Funkhouser. 2023. TidyBot: Personalized Robot Assistance with Large Language Models. *Autonomous Robots* (2023). arXiv:2305.05658.
 - [266] Philipp Wu, Alejandro Escontrela, Danijar Hafner, Ken Goldberg, and Pieter Abbeel. 2022. DayDreamer: World Models for Physical Robot Learning. In *Conference on Robot Learning (CoRL)*. arXiv:2206.14176.
 - [267] Ruihai Wu, Yan Zhao, Kaichun Mo, Zizheng Guo, Yian Wang, et al. 2022. VAT-Mart: Learning Visual Action Trajectory Proposals for Manipulating 3D Articulated Objects. In *International Conference on Learning Representations (ICLR)*. arXiv:2106.14440.
 - [268] Yue Wu, Yewen Fan, Paul Pu Liang, Amos Azaria, Yuanzhi Li, and Tom M. Mitchell. 2023. Read and Reap the Rewards: Learning to Play Atari with the Help of Instruction Manuals. *arXiv preprint arXiv:2302.04449* (2023).
 - [269] Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, et al. 2020. SAPIEN: A SimulAteD Part-based Interactive ENvironment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:2003.08515.
 - [270] Ted Xiao, Harris Chan, Pierre Sermanet, Ayzaan Wahid, Anthony Brohan, Karol Hausman, Sergey Levine, and Jonathan Tompson. 2022. Robotic Skill Acquisition via Instruction Augmentation with Vision-Language Models. *arXiv preprint arXiv:2211.11736* (2022).
 - [271] Tete Xiao, Ilija Radosavovic, Trevor Darrell, and Jitendra Malik. 2022. Masked Visual Pre-training for Motor Control. *arXiv preprint arXiv:2203.06173* (2022).
 - [272] Wenli Xiao, Jia Xie, Tonghe Zhang, Haotian Lin, Letian Fu, Haoru Xue, Jalen Lu, Yi Yang, Cunxi Dai, Zi Wang, Jimmy Wu, Guanzhi Wang, S. Shankar Sastry, Ken Goldberg, Linxi Fan, Yuke Zhu, and Guanya Shi. 2026. ENPIRE: Agentic Robot Policy Self-Improvement in the Real World. *arXiv preprint* (2026). arXiv:2606.19980.
 - [273] Senwei Xie, Hongyu Wang, Zhanqi Xiao, Ruiping Wang, and Xilin Chen. 2025. Robotic Programmer: Video Instructed Policy Code Generation for Robotic Manipulation. *arXiv preprint* (2025). arXiv:2501.04268.
 - [274] Senwei Xie, Yuntian Zhang, Ruiping Wang, and Xilin Chen. 2026. Uni-Skill: Building Self-Evolving Skill Repository for Generalizable Robotic Manipulation. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2603.02623.
 - [275] Tianbao Xie, Siheng Zhao, Chen Henry Wu, Yitao Liu, Qian Luo, Victor Zhong, Yanchao Yang, and Tao Yu. 2024. Text2Reward: Reward Shaping with Language Models for Reinforcement Learning. In *International Conference on Learning Representations (ICLR)*. arXiv:2309.11489.
 - [276] Jianwei Yang et al. 2025. Magma: A Foundation Model for Multimodal AI Agents. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:2502.13130.
 - [277] Mengjiao Yang, Yilun Du, Kamyar Ghasemipour, Jonathan Tompson, Dale Schuurmans, and Pieter Abbeel. 2024. Learning Interactive Real-World Simulators. In *International Conference on Learning Representations (ICLR)*.

- arXiv:2310.06114.
- [278] Rushuai Yang, Chenjia Bai, Hongyi Guo, Siyuan Li, Bin Zhao, Zhen Wang, Peng Liu, and Xuelong Li. 2023. Behavior Contrastive Learning for Unsupervised Skill Discovery. In *International Conference on Machine Learning (ICML)*. arXiv:2305.04477.
 - [279] Sizhe Yang, Qian Luo, Anupam Pani, and Yanchao Yang. 2024. BBSEA: An Exploration of Brain-Body Synchronization for Embodied Agents. *arXiv preprint arXiv:2402.08212* (2024).
 - [280] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. ReAct: Synergizing Reasoning and Acting in Language Models. *arXiv preprint arXiv:2210.03629* (2022).
 - [281] Denis Yarats, Rob Fergus, Alessandro Lazaric, and Lerrel Pinto. 2021. Reinforcement Learning with Prototypical Representations. In *International Conference on Machine Learning (ICML)*. arXiv:2102.11271.
 - [282] Seonghyeon Ye et al. 2025. Latent Action Pretraining from Videos. In *International Conference on Learning Representations (ICLR)*. arXiv:2410.11758.
 - [283] Weirui Ye, Shaohuai Liu, Thanard Kurutach, Pieter Abbeel, and Yang Gao. 2021. Mastering Atari Games with Limited Data. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2111.00210.
 - [284] Naoki Yokoyama, Sehoon Ha, Dhruv Batra, Jiuguang Wang, and Bernadette Bucher. 2024. VLFM: Vision-Language Frontier Maps for Zero-Shot Semantic Navigation. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2312.03275.
 - [285] Takuma Yoneda, Jiading Fang, Peng Li, Huanyu Zhang, Tianchong Jiang, Shengjie Lin, Ben Picker, David Yunis, Hongyuan Mei, and Matthew R. Walter. 2024. Statler: State-Maintaining Language Models for Embodied Reasoning. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2306.17840.
 - [286] Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Avnish Narayan, Hayden Shively, Adithya Bellathur, Karol Hausman, Chelsea Finn, and Sergey Levine. 2019. Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning. In *Conference on Robot Learning (CoRL)*. arXiv:1910.10897.
 - [287] Tianhe Yu, Ted Xiao, Austin Stone, Jonathan Tompson, Anthony Brohan, et al. 2023. Scaling Robot Learning with Semantically Imagined Experience. *arXiv preprint arXiv:2302.11550* (2023).
 - [288] Wenhao Yu, Nimrod Gileadi, Chuyuan Fu, Sean Kirmani, Kuang-Huei Lee, Montse Gonzalez Arenas, Hao-Tien Lewis Chiang, Tom Erez, Leonard Hasenclever, Jan Humplik, Brian Ichter, Ted Xiao, Peng Xu, Andy Zeng, Tingnan Zhang, Nicolas Heess, Dorsa Sadigh, Jie Tan, Yuval Tassa, and Fei Xia. 2023. Language to Rewards for Robotic Skill Synthesis. In *Conference on Robot Learning (CoRL)*. arXiv:2306.08647.
 - [289] Wentao Yuan et al. 2024. RoboPoint: A Vision-Language Model for Spatial Affordance Prediction for Robotics. *arXiv preprint arXiv:2406.10721* (2024).
 - [290] Michał Zawalski et al. 2024. Robotic Control via Embodied Chain-of-Thought Reasoning. *arXiv preprint arXiv:2407.08693* (2024).
 - [291] Yanjie Ze et al. 2024. 3D Diffusion Policy: Generalizable Visuomotor Policy Learning via Simple 3D Representations. In *Robotics: Science and Systems (RSS)*. arXiv:2403.03954.
 - [292] Andy Zeng et al. 2020. Transporter Networks: Rearranging the Visual World for Robotic Manipulation. In *Conference on Robot Learning (CoRL)*. arXiv:2010.14406.
 - [293] Andy Zeng, Maria Attarian, Brian Ichter, Krzysztof Choromanski, Adrian Wong, et al. 2022. Socratic Models: Composing Zero-Shot Multimodal Reasoning with Language. *arXiv preprint arXiv:2204.00598* (2022).
 - [294] Lihan Zha, Yuchen Cui, Li-Heng Lin, Minae Kwon, Montserrat Gonzalez Arenas, Andy Zeng, Fei Xia, and Dorsa Sadigh. 2024. Distilling and Retrieving Generalizable Knowledge for Robot Manipulation via Language Corrections. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2311.10678.
 - [295] Jesse Zhang, Karl Pertsch, Jiahui Zhang, and Joseph J. Lim. 2024. SPRINT: Scalable Policy Pre-Training via Language Instruction Relabeling. In *IEEE International Conference on Robotics and Automation (ICRA)*. arXiv:2306.11886.
 - [296] Jesse Zhang, Haonan Yu, and Wei Xu. 2021. Hierarchical Reinforcement Learning By Discovering Intrinsic Options. In *International Conference on Learning Representations (ICLR)*. arXiv:2101.06521.
 - [297] Jesse Zhang, Jiahui Zhang, Karl Pertsch, Ziyi Liu, Xiang Ren, Minsuk Chang, Shao-Hua Sun, and Joseph J. Lim. 2023. Bootstrap Your Own Skills: Learning to Solve New Tasks with Large Language Model Guidance. In *Conference on Robot Learning (CoRL)*. arXiv:2310.10021.
 - [298] Weipu Zhang, Gang Wang, Jian Sun, Yetian Yuan, and Gao Huang. 2023. STORM: Efficient Stochastic Transformer based World Models for Reinforcement Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv:2310.09615.
 - [299] Yang Zhang, Shixin Yang, Chenjia Bai, Fei Wu, Xiu Li, Zhen Wang, and Xuelong Li. 2024. Towards Efficient LLM Grounding for Embodied Multi-Agent Collaboration. *arXiv preprint arXiv:2405.14314* (2024).
 - [300] Ziao Zhang, Kou Shi, Shiting Huang, Avery Nie, Yu Zeng, Yiming Zhao, Zhen Fang, Qisheng Su, Haibo Qiu, Wei Yang, Qingnan Ren, Shun Zou, Wenxuan Huang, Lin Chen, Zehui Chen, and Feng Zhao. 2026. SkillFlow: Benchmarking Lifelong Skill Discovery and Evolution for Autonomous Agents. *arXiv preprint* (2026). arXiv:2604.17308.

- [301] Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. 2024. ExpeL: LLM Agents Are Experiential Learners. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. arXiv:2308.10144.
- [302] Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. 2023. Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware. In *Robotics: Science and Systems (RSS)*. arXiv:2304.13705.
- [303] Haoyu Zhen et al. 2024. 3D-VLA: A 3D Vision-Language-Action Generative World Model. In *International Conference on Machine Learning (ICML)*. arXiv:2403.09631.
- [304] Ruijie Zheng et al. 2024. TraceVLA: Visual Trace Prompting Enhances Spatial-Temporal Awareness for Generalist Robotic Policies. *arXiv preprint arXiv:2412.10345* (2024).
- [305] Ying Zheng et al. 2025. A Survey of Embodied Learning for Object-Centric Robotic Manipulation. *Machine Intelligence Research* (2025). arXiv:2408.11537.
- [306] Enshen Zhou, Qi Su, Cheng Chi, Zhizheng Zhang, Zhongyuan Wang, Tiejun Huang, Lu Sheng, and He Wang. 2025. Code-as-Monitor: Constraint-aware Visual Programming for Reactive and Proactive Robotic Failure Detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. arXiv:2412.04455.
- [307] Siyuan Zhou, Yilun Du, Jiaben Chen, Yandong Li, Dit-Yan Yeung, and Chuang Gan. 2024. RoboDreamer: Learning Compositional World Models for Robot Imagination. *arXiv preprint arXiv:2404.12377* (2024).
- [308] Zhehua Zhou, Jiayang Song, Kunpeng Yao, Zhan Shu, and Lei Ma. 2023. ISR-LLM: Iterative Self-Refined Large Language Model for Long-Horizon Sequential Task Planning. *arXiv preprint arXiv:2308.13724* (2023).
- [309] Xizhou Zhu, Yuntao Chen, Hao Tian, Chenxin Tao, Weijie Su, Chenyu Yang, Gao Huang, Bin Li, Lewei Lu, Xiaogang Wang, Yu Qiao, Zhaoxiang Zhang, and Jifeng Dai. 2023. Ghost in the Minecraft: Generally Capable Agents for Open-World Environments via Large Language Models with Text-based Knowledge and Memory. *arXiv preprint* (2023). arXiv:2305.17144.
- [310] Yuke Zhu, Josiah Wong, Ajay Mandlekar, Roberto Martín-Martín, Abhishek Joshi, Kevin Lin, Abhiram Maddukuri, Soroush Nasiriany, and Yifeng Zhu. 2020. robosuite: A Modular Simulation Framework and Benchmark for Robot Learning. *arXiv preprint* (2020). arXiv:2009.12293.
- [311] Ziwen Zhuang, Shenzhe Yao, and Hang Zhao. 2024. Humanoid Parkour Learning. In *Conference on Robot Learning (CoRL)*. arXiv:2406.10759.

A THE BROADER ROBOT-LEARNING LANDSCAPE

The taxonomy of Figure 2 foregrounds 77 systems chosen, by the placement criteria of §2.1, to populate a single analytical axis, weights versus skills, rather than to enumerate the field. To situate those systems within the wider research landscape, and to make the survey’s coverage inspectable at a glance, Table 11 catalogues a further set of representative works: end-to-end and generalist policies, imitation and diffusion policies, language-model planning and task-and-motion planning, reward and data synthesis, unsupervised skill discovery and hierarchical reinforcement learning, world models and model-based control, representation learning and offline reinforcement learning, manipulation, dexterity, and locomotion, and the datasets, benchmarks, and simulators that support them. Together with the 77 systems detailed in the main text (Tables 2–6), the works surveyed number more than three hundred. The list is representative rather than exhaustive; each entry is a real, individually citeable paper, grouped by area and ordered by year. This set is characterised in the main text (§2.2): Figure 5 summarises the corpus at a glance, Figure 6 profiles it by publication year, evaluation domain, learning signal, and embodiment, and the visual galleries (Figures 8 and 9) show the surveyed systems; Table 10 lists, for each gallery panel, the source paper and the figure number within it, so that no image is used without a traceable, clickable origin.

Table 11. **The broader robot-learning landscape: 225 representative works, compared on six axes.** Beyond the 77 systems placed in the taxonomy (Figure 2, Tables 2–6), this table catalogues the wider field, grouped into eleven areas and sub-grouped by family. Columns: **Year**, **Ships** (what the method contributes), **Learn** (primary learning signal: IL, RL, offline-RL, LLM, self-sup., model-based), **Eval** (real/sim/game/offline), **Embod.** (embodiment), and run-time **S.I.** (self-improvement: ✓ yes, ~ partial, ✗ no). A – marks not-applicable or unverified. Every entry is a real, individually citeable paper.

System / work	Year	Ships	Learn	Eval	Embod.	S.I.
End-to-end & generalist policies (28)						
<i>Vision-language-action</i>						
3D-VLA [303]	2024	world-model	LLM	sim	arm	✗
ECot [290]	2024	policy	IL	real	arm	✗
LLaRA [140]	2024	policy	LLM	sim	arm	✗
ManipLLM [141]	2024	policy	LLM	sim	arm	✗
RoboMamba [154]	2024	policy	LLM	sim	arm	✗
RoboPoint [289]	2024	plan	LLM	real	multi	✗
TinyVLA [262]	2024	policy	IL	real	arm	✗
TraceVLA [304]	2024	policy	IL	sim	arm	✗
DexVLA [263]	2025	policy	IL	real	multi	✗
FAST [205]	2025	repr	IL	real	arm	✗
OpenVLA-OFT [119]	2025	policy	IL	sim	arm	✗
<i>Generalist / multi-task</i>						
QT-Opt [113]	2018	policy	RL	real	arm	~
MT-Opt [114]	2021	policy	RL	real	arm	~
Gato [211]	2022	policy	IL	sim	multi	✗
PaLM-E [45]	2023	plan	LLM	real	multi	✗
Q-Transformer [23]	2023	policy	offline-RL	real	arm	✗
UniPi [46]	2023	policy	model-based	sim	arm	✗
VIMA [110]	2023	policy	IL	sim	arm	✗
GR-2 [22]	2024	policy	IL	real	arm	✗
LEO [96]	2024	policy	LLM	sim	multi	✗
Magma [276]	2025	weights	LLM	sim	arm	✗
<i>Video / action pretraining</i>						
ATM [261]	2024	policy	self-sup	sim	arm	✗
GR-1 [264]	2024	policy	IL	sim	arm	✗
HPT [252]	2024	repr	IL	sim	multi	✗
VPP [94]	2024	policy	IL	sim	arm	✗
Cosmos [193]	2025	world-model	self-sup	–	–	✗
LAPA [282]	2025	policy	self-sup	sim	arm	✗

Table 11 (continued)

System / work	Year	Ships	Learn	Eval	Embod.	S.I.
<i>LLM reward synthesis</i>						
ELLM [47]	2023	reward	RL	game	—	✗
LAMP [2]	2023	reward	RL	sim	arm	✗
Motif [121]	2023	reward	RL	game	—	✗
Read and Reap [268]	2023	reward	RL	game	—	✗
VLM-RM [216]	2023	reward	RL	sim	humanoid	✗
Code as Reward [245]	2024	reward	RL	sim	—	✗
RL-VLM-F [255]	2024	reward	RL	sim	arm	✗
<i>Generative data / augmentation</i>						
CACTI [171]	2022	data	IL	real	arm	✗
DIAL [270]	2022	data	IL	real	mobile	✗
GenAug [34]	2023	data	IL	real	arm	✗
GenSim [253]	2023	data	IL	sim	arm	✗
ROSIE [287]	2023	data	IL	real	mobile	✗
Scaling Up Distilling Down [78]	2023	policy	IL	real	arm	✗
BBSEA [279]	2024	skill	IL	sim	arm	~
DexMimicGen [111]	2024	data	IL	sim	humanoid	✗
GenSim2 [95]	2024	data	IL	sim	arm	✗
RoboDreamer [307]	2024	world-model	model-based	—	arm	✗
Unsupervised skill discovery & hierarchical RL (25)						
<i>Unsupervised skill discovery</i>						
VIC [70]	2016	skill	RL	sim	—	✗
Eigenoption [166]	2018	skill	RL	game	—	✗
VALOR [1]	2018	skill	RL	sim	—	✗
EDL [20]	2020	skill	RL	sim	—	✗
VISR [87]	2020	skill	RL	game	—	~
APS [151]	2021	skill	RL	game	—	✗
APT [152]	2021	skill	RL	game	—	✗
ProtoRL [281]	2021	repr	self-sup	sim	—	✗
WURL [88]	2022	skill	RL	sim	—	✗
BeCL [278]	2023	skill	RL	sim	—	✗
Choreographer [176]	2023	skill	model-based	sim	—	✗
CSD [201]	2023	skill	RL	sim	multi	✗
DGPO [31]	2024	skill	RL	sim	—	✗
<i>Hierarchical / option RL</i>						
h-DQN [124]	2016	policy	RL	game	—	✗
FeUdal Networks [246]	2017	policy	RL	game	—	✗
Option-Critic [10]	2017	skill	RL	game	—	✗
SNN4HRL [55]	2017	skill	RL	sim	multi	✗
HIRO [187]	2018	policy	RL	sim	legged	✗
MLSH [56]	2018	skill	RL	sim	multi	~
HAC [132]	2019	policy	RL	sim	arm	✗
MCP [204]	2019	skill	RL	sim	humanoid	✗
NPMP [180]	2019	skill	IL	sim	humanoid	✗
Relay Policy Learning [74]	2019	policy	IL	sim	arm	✗
HIDIO [296]	2021	skill	RL	sim	multi	✗
Director [79]	2022	policy	model-based	sim	—	✗
World models & model-based RL (24)						
<i>World models</i>						
World Models [77]	2018	world-model	model-based	game	—	✗
PlaNet [81]	2019	world-model	model-based	sim	—	✗
Dreamer [80]	2020	world-model	model-based	sim	—	✗
DreamerV2 [82]	2021	world-model	model-based	game	—	✗
DayDreamer [266]	2022	world-model	model-based	real	multi	✓
Iso-Dream [198]	2022	world-model	model-based	sim	—	✗
MWM [221]	2022	world-model	model-based	sim	arm	✗
TransDreamer [25]	2022	world-model	model-based	game	—	✗
DreamerV3 [83]	2023	world-model	model-based	game	—	✗
IRIS [181]	2023	world-model	model-based	game	—	✗
STORM [298]	2023	world-model	model-based	game	—	✗
SWIM [179]	2023	world-model	model-based	real	arm	✗

Table 11 (continued)

System / work	Year	Ships	Learn	Eval	Embod.	S.I.
TWM [215]	2023	world-model	model-based	game	–	✗
DIAMOND [8]	2024	world-model	model-based	game	–	✗
Genie [19]	2024	world-model	self-sup	game	–	✗
UniSim [277]	2024	world-model	self-sup	real	arm	✗
<i>Model-based RL</i>						
PETS [39]	2018	world-model	model-based	sim	–	✗
MBPO [108]	2019	policy	model-based	sim	–	✗
MuZero [220]	2020	world-model	model-based	game	–	✗
SimPLe [112]	2020	world-model	model-based	game	–	✗
SLAC [130]	2020	repr	RL	sim	–	✗
EfficientZero [283]	2021	world-model	model-based	game	–	✗
TD-MPC [86]	2022	world-model	model-based	sim	–	✗
TD-MPC2 [85]	2024	world-model	model-based	sim	multi	✗
Representation learning & offline RL (17)						
<i>Visual representation learning</i>						
MVP [271]	2022	repr	self-sup	sim	arm	✗
PVR [199]	2022	repr	self-sup	sim	multi	✗
R3M [189]	2022	repr	self-sup	sim	arm	✗
LIV [162]	2023	reward	self-sup	real	arm	✗
VC-1 [169]	2023	repr	self-sup	sim	multi	✗
VIP [165]	2023	reward	self-sup	real	arm	✗
Voltron [116]	2023	repr	self-sup	sim	arm	✗
Theia [227]	2024	repr	self-sup	sim	arm	✗
<i>Offline RL</i>						
BCQ [62]	2019	policy	offline-RL	offline	–	✗
AWAC [188]	2020	policy	offline-RL	offline	–	~
CQL [126]	2020	policy	offline-RL	offline	–	✗
Decision Transformer [28]	2021	policy	offline-RL	offline	–	✗
TD3+BC [61]	2021	policy	offline-RL	offline	–	✗
Trajectory Transformer [109]	2021	policy	offline-RL	offline	–	✗
IQL [123]	2022	policy	offline-RL	offline	–	✗
Cal-QL [190]	2023	policy	offline-RL	offline	–	~
Diffusion-QL [260]	2023	policy	offline-RL	offline	–	✗
Manipulation, dexterity & sim-to-real (16)						
<i>Manipulation / grasping</i>						
Dex-Net 2.0 [168]	2017	policy	–	real	arm	✗
GraspNet-1Billion [53]	2020	bench	–	real	arm	✗
Transporter Nets [292]	2020	policy	IL	sim	arm	✗
CLIPort [233]	2021	policy	IL	sim	arm	✗
Contact-GraspNet [239]	2021	code	–	real	arm	✗
AnyGrasp [52]	2023	code	–	real	arm	✗
<i>Dexterous / in-hand</i>						
Dactyl [9]	2018	policy	RL	real	hand	~
DAPG [209]	2018	policy	RL	sim	hand	✗
Rubiks Cube Hand [7]	2019	policy	RL	real	hand	~
DexPilot [84]	2020	code	–	real	hand	✗
Visual Dexterity [30]	2023	policy	RL	real	hand	✗
<i>Sim-to-real transfer</i>						
CAD2RL [218]	2017	policy	RL	real	mobile	✗
Domain Randomization [241]	2017	code	–	real	arm	✗
Dynamics Randomization [203]	2018	policy	RL	real	arm	~
RCAN [104]	2019	policy	RL	real	arm	✗
SimOpt [24]	2019	policy	RL	real	arm	~
Locomotion & humanoid control (16)						
<i>Legged locomotion</i>						
ANYmal [103]	2019	policy	RL	real	legged	✗
Quadruped Terrain [131]	2020	policy	RL	real	legged	~
Legged Gym [217]	2021	code	RL	real	legged	✗
RMA [125]	2021	policy	RL	real	legged	~
Egocentric Locomotion [3]	2022	policy	RL	real	legged	~

Table 11 (continued)

System / work	Year	Ships	Learn	Eval	Embod.	S.I.
Perceptive Locomotion [182]	2022	policy	RL	real	legged	~
Walk These Ways [175]	2022	policy	RL	real	legged	✗
Extreme Parkour [36]	2024	policy	RL	real	legged	✗
<i>Humanoid control</i>						
ExBody [35]	2024	policy	RL	real	humanoid	✗
H2O [91]	2024	policy	RL	real	humanoid	✗
Humanoid Locomotion [208]	2024	policy	RL	real	humanoid	~
Humanoid Parkour [311]	2024	policy	RL	real	humanoid	✗
HumanPlus [59]	2024	policy	IL	real	humanoid	✗
OmniH2O [90]	2024	policy	RL	real	humanoid	✗
ASAP [89]	2025	policy	RL	real	humanoid	~
HOVER [92]	2025	policy	RL	real	humanoid	✗
Navigation, tactile & affordance (15)						
<i>Visual / language navigation</i>						
LM-Nav [224]	2022	plan	LLM	real	mobile	✗
ViKiNG [223]	2022	policy	IL	real	mobile	✗
CoW [63]	2023	bench	–	sim	mobile	✗
GNM [225]	2023	policy	IL	real	multi	✗
ViNT [226]	2023	policy	IL	real	mobile	✗
NoMaD [238]	2024	policy	IL	real	mobile	✗
VLFM [284]	2024	policy	LLM	sim	mobile	✗
<i>Tactile & affordance</i>						
DIGIT [128]	2020	–	–	real	hand	✗
Where2Act [184]	2021	repr	self-sup	sim	arm	✗
FlowBot3D [49]	2022	policy	self-sup	sim	arm	✗
HULC [177]	2022	policy	IL	sim	arm	✗
VAT-Mart [267]	2022	plan	RL	sim	arm	✗
See-to-Touch [75]	2023	policy	RL	real	hand	~
T-DEX [76]	2023	repr	self-sup	real	hand	✗
VRB [11]	2023	repr	self-sup	real	arm	✗
Datasets, benchmarks & simulators (16)						
<i>Datasets</i>						
RoboNet [40]	2019	data	model-based	real	arm	✗
Ego4D [69]	2022	data	–	offline	–	✗
BridgeData V2 [247]	2023	data	IL	real	arm	✗
DobbE [222]	2023	data	IL	real	mobile	✗
RH20T [51]	2023	data	IL	real	arm	✗
DROID [118]	2024	data	IL	real	arm	✗
<i>Benchmarks</i>						
D4RL [57]	2020	bench	offline-RL	offline	–	✗
ManiSkill [185]	2021	bench	IL	sim	arm	✗
<i>Simulators</i>						
AI2-THOR [122]	2017	code	–	sim	mobile	✗
Habitat [219]	2019	code	–	sim	mobile	✗
SAPIEN [269]	2020	code	–	sim	arm	✗
ThreeDWorld [64]	2020	code	–	sim	multi	✗
iGibson [230]	2021	code	–	sim	mobile	✗
Isaac Gym [170]	2021	code	–	sim	multi	✗
Isaac Lab Orbit [183]	2023	code	–	sim	multi	✗
RoboCasa [191]	2024	bench	IL	sim	mobile	✗

Table 10. **Per-panel provenance for Figures 8, 9, and 1.** Every panel is cropped from the listed figure of the cited paper; the same links are embedded, clickable, in the figures themselves. † marks a panel whose source figure is a redrawn/replicated variant or appears identically more than once in the paper (matched by content). The corpus-at-a-glance collage (Figure 1) reuses the robot panels and is not relisted.

System	Paper	Source figure	Source file
<i>Figure 8: robots (27 panels)</i>			
RT-1	arXiv:2212.06817	Fig. 1	rt1_teaser_tasks.png
RT-2	arXiv:2307.15818	Fig. 2	RT2-capabilities-dm.png
OpenVLA	arXiv:2406.09246	Fig. 11	droid_wipe_task.jpeg
Octo	arXiv:2405.12213	Fig. 4 †	pdf_im-164.png
π_0	arXiv:2410.24164	Fig. 2	fig2_final.jpeg
CogACT	arXiv:2411.19650	Fig. II (suppl.)	franka_robot_with_label.png
ProgPrompt	arXiv:2209.11302	Fig. 1	pdf_im-000.png
VoxPoser	arXiv:2307.05973	Fig. 1	pdf_im-005.png
Code-as-Monitor	arXiv:2412.04455	Fig. 1	pdf_im-002.png
REFLECT	arXiv:2306.15724	Fig. 1	pdf_im-029.png
LOTUS	arXiv:2311.02058	Fig. 1	pdf_im-014.png
CaP-X	arXiv:2603.22435	Fig. 21	seq4.png
UniSkill	arXiv:2603.02623	Fig. 4	pdf_im-121.png
Gemini Robotics	arXiv:2503.20020	Fig. 10	mv1.jpeg
TidyBot	arXiv:2305.05658	Fig. 4	IMG_5028.jpg
DrEureka	arXiv:2406.01967	Fig. 4	quadruped_terrains.png
DrEureka	arXiv:2406.01967	Fig. 2	robots.png
ENPIRE	arXiv:2606.19980	Fig. 11	four_camera_setup.jpg
CrossFormer	arXiv:2408.11812	Fig. 4	pdf_im-099.png
CrossFormer	arXiv:2408.11812	Fig. 4	pdf_im-098.png
Meta-World	arXiv:1910.10897	Fig. 1	pdf_im-000.png
robosuite	arXiv:2009.12293	env fig. (unnum., p.14) †	env_door_v15.png
Instruct2Act	arXiv:2305.11176	Fig. 1	pdf_im-003.png
Prompt2Walk	arXiv:2309.09969	Fig. 1	pdf_im-062.png
DoReMi	arXiv:2307.00329v3	Fig. 11b	ft1.png
DIAYN	arXiv:1802.06070	Fig. 3	skills.jpg
Mirage	arXiv:2402.19249	Fig. 2	Sim_Tasks_Robots_Fig.png
<i>Figure 9: results (20 panels)</i>			
RT-2	arXiv:2307.15818	Fig. 6a	rt2_emergent_dm.png
Code-as-Policies	arXiv:2209.07753	Fig. 4	generalization_types_flat.png
Demo2Code	arXiv:2305.16744	Fig. 13	code_ablation_v4.png
Eureka	arXiv:2310.12931	Fig. 13	dexterity_bar_chart.png
RoboScript	arXiv:2402.14623	Fig. 8	ablation_drawer_place.png
π_0	arXiv:2410.24164	Fig. 13	complex_finetune.png
Text2Motion	arXiv:2303.12153	Fig. 5	Figure_5.jpg
DrEureka	arXiv:2406.01967	Fig. 8	quadruped_training_curves.png
Eureka	arXiv:2310.12931	Fig. 10	bidex_training_curves.png
ManiSkill2	arXiv:2302.04659	Fig. 3	CNNE.png
$\pi_{0.5}$	arXiv:2504.16054	Fig. 9	env_scaling_results0.png
Voyager	arXiv:2305.16291	Fig. A.4	model_variations.png
DIAYN	arXiv:1802.06070	Fig. 4	cheetah_entropy.png
DADS	arXiv:1907.01657	Fig. 8	hierarchical_control_usl_fix4.png
π_0	arXiv:2410.24164	Fig. 4	combined-robot-allocation-chart.png
SkiMo	arXiv:2207.07560	Fig. 9	ours_coverage.png
OPAL	arXiv:2010.13611	Fig. 1	antmaze_medium.png
LRL	arXiv:2406.18746	Fig. 3 †	tsne231.drawio.png
LSD	arXiv:2202.00914	Fig. 1b †	pdf_im-001.png
OPAL	arXiv:2010.13611	Fig. 3	cql_large_heat.png