

Intelligence Across Embodiments

Bo Ai^{1,2}, Henrik I. Christensen², and Hao Su³

¹ Stanford University

² University of California San Diego

³ Sudo AI GmbH

Abstract. Robotic embodiment encompasses the sensing, kinematics, dynamics, geometry, actuation, and control through which an agent physically interacts with the world. These properties vary across robots and change over time. We argue that general embodied intelligence requires learning that accumulates across these differences. Prevailing methods that engineer correspondences to bridge embodiment differences offer immediate practical gains, but their assumptions limit the scope of transfer in the long run. Instead, a more general approach should discover representations that support transfer to a larger range of embodiments as experience grows. We propose embodiment diversity as a promising axis of scaling, and identify broad learned priors as a complementary ingredient. We call for evaluations that better characterize embodiment gaps and transfer performance. More broadly, cross-embodiment learning connects the practical challenge of learning from heterogeneous robot experience with a broader scientific pursuit inspired by nature—physical intelligence that adapts and co-evolves with its embodiments to gain agency over its behavior and physical forms.

Keywords: robotics, embodied intelligence

Experience Beyond One Embodiment

Robotics has long sought machines capable of autonomously performing physical tasks in diverse environments. Traditionally, perception, planning, and control have provided computational foundations for specifying behavior through models and algorithms [29, 38, 49]. Machine learning offers a complementary approach, acquiring sensorimotor mappings from data and enabling capabilities difficult to specify manually [2, 24, 28, 30]. Extending these successes to broadly capable robots depends on the experience available for learning. A widely discussed obstacle is the scarcity of robot data [3, 17].

Learning systems generally improve from two sources of experience: data generated by other agents, often humans, and data generated through their own interactions with an environment. The former can bootstrap the latter. Language models, for example, pre-train on human-generated text and subsequently improve specific capabilities, such as programming, through reinforcement learning [4, 39]; AlphaGo combined learning from human games with self-play [45].

Robot learning can draw on both sources in principle, but its experience is mediated by embodiment—the configuration of sensing, morphology, kinematics, dynamics, actuation, and low-level control through which an agent interacts with the world. Embodiment determines what a robot can perceive and do, which strategies it can realize, and how its interactions are manifested as data. Unlike language models that see the same languages from different texts, robot policies see distinct sensorimotor “vocabularies” given by different embodiments.

This heterogeneity is intrinsic to embodied agents. Different goals and environments demand different physical interactions, which favor different embodiments. Embodiment variation occurs along two axes: across populations and over time. Across populations, biological evolution has produced diverse embodiments adapted to different ecological niches and ways of life [18, 36]. For robot design, factories, homes, disaster sites, and extraterrestrial environments impose different requirements on geometry, mobility, precision, strength, compliance, and sensing [26]. Even within the same design, variation persists because of manufacturing tolerances, payloads, calibration, sensors, and controller configurations. Across time, actuators wear, sensors drift, controllers are recalibrated, and successive hardware generations are redesigned manually or computationally [18, 19, 57]. Thus, a general learning system needs to span both populations and generations of embodiments. Intelligence is the persistent entity in this continual learning process that spans changing physical entities.

From Engineered Alignment to Learned Representations

A common strategy for learning and transfer across embodiments is to engineer correspondences. At the hardware level, standardizing gripper designs aligns handheld data collection devices with the target end effector [9, 13, 54], and low-cost leader arms reproduce the target robot’s kinematics for joint-space teleoperation [15, 58]. Anthropomorphic hands and humanoids exploit structural similarities with humans to facilitate motion retargeting [14, 20, 22, 51]. At the level of model representations, image goals and flow predictions describe desired visual or geometric changes that can be shared across embodiments [21, 31, 42, 53]. Interaction intent and contact provide interfaces for relating physical interactions across embodiments [43, 50, 52], and end-effector actions abstract away joint-level differences [9, 21, 55]. These approaches establish shared interfaces with human knowledge to facilitate learning and transfer.

Many of these approaches rely on human-designed priors whose assumptions about tasks and environments can become long-term bottlenecks. An end-effector action space assumes that specifying end-effector motion is sufficient for task completion, which becomes limiting when tasks demand whole-body coordination or particular dynamic responses. Contact-based correspondence presumes that the target embodiment can reproduce the relevant contacts within its geometric, kinematic, and dynamic constraints. When these assumptions fail, successful execution may require a different interaction strategy that the prescribed interface cannot express. Humanoid designs draw on humans as an existence

proof of broad physical capability, but the diversity of embodiments in nature suggests that the human form need not be best suited to every physical task and environment [18, 36]. These human-designed priors enable short-term progress, but their assumptions can constrain the improvement that we would like to see from increasing experience and computation.

The more fundamental question is whether we can build systems that *discover* the general representations needed for transfer through learning. A recurring pattern in artificial intelligence research is that representations learned with increasing data and computation can surpass human-designed priors [48]. Computer vision moved from hand-designed features toward representations learned by neural networks [27, 37]. Chess systems combined search with policies and value functions learned through self-play, reducing reliance on manually designed game features [46]. Analogously, cross-embodiment learning may discover abstractions ranging from embodiment-agnostic understanding to embodiment-specific predictions.

Such abstractions might capture task semantics (the intended outcomes of a command), spatial relations (the relative positions and orientations of objects and the agent), physical interactions (the consequences of actions), and control (joint coordination that realizes desired behavior). These categories describe one possible organization, and prior work explicitly models some of them. Learned representations need not follow this hierarchy and may instead encode invariances or relations that vary systematically with embodiment, such as latent action representations that associate different motor commands across embodiments with similar physical effects. The goal is for accumulating experience to improve these representations, broaden transfer, and reduce adaptation costs on new embodiments. The scientific question is what ingredients can support this cumulative improvement.

What Enables Cumulative Transfer?

Broader pretraining can make subsequent experience more useful. Human demonstrations produce greater gains after broad robot pretraining [23], while scaling pretraining on human video improves performance after robot-specific fine-tuning [12]. These findings suggest that the value of new data depends partly on the learner’s prior representations. Humans offer an intuitive analogy: a pilot’s understanding of an aircraft helps connect simulated experience to the controls, dynamics, and consequences of physical flight. Broad learned priors may similarly help robots interpret experience acquired through other embodiments and ground it in their own sensorimotor interfaces.

Embodiment diversity is a plausible source of such priors and should be treated as a primary axis of scaling alongside tasks and environments. Broader task pretraining can improve adaptation to new tasks [5], while data collected across environments supports generalization beyond the settings in which it was collected [25]. Experience across embodiments may similarly reveal common interaction structure and how successful behavior changes with physical capabil-

ities. Early results support this direction. Training across diverse end effectors enables adaptation to new hands and tools [16], while simulation-based training supports zero-shot transfer to unseen embodiments in locomotion and dexterous manipulation [1, 41]. In locomotion, increasing embodiment diversity has also produced broader generalization than increasing data on fixed embodiments [1]. Establishing how far these gains extend requires systematic variation in embodiment alongside tasks and environments.

Obtaining such experience poses practical challenges. Collecting diverse real-world embodiment data requires constructing, maintaining, and calibrating different physical systems, as well as collecting expert demonstrations on each. Simulation offers a practical platform for synthesizing diverse embodiments, varying their properties systematically, and studying transfer at scale. Generating embodiments requires mechanically feasible morphologies with compatible dynamics and actuation properties, together with controller settings that produce a range of closed-loop responses. Covering the range of embodiments encountered in practice therefore requires an understanding of the dependencies among geometry, mass distribution, actuator capabilities, and control. Generating demonstrations presents another challenge: planning can be computationally expensive, while reinforcement learning may require substantial reward engineering. Furthermore, simulated experience must improve learning and performance on real robots. Bridging the simulation-to-reality gap remains an open challenge, but recent results with visuomotor policies trained on simulated data show signs of promise [11].

Training policies across a much larger population of embodiments raises fundamental questions about model architecture. Existing datasets typically contain on the order of 10^2 embodiments, but systematic variation could expand this number by orders of magnitude. How should a model learn jointly from distinct observation and action spaces? What information about embodiment should be supplied explicitly, and what can be inferred through interaction? Models that rely on visual observations to identify embodiment may struggle as diversity grows, since visually similar robots can differ in dynamics and controller response. This motivates architectures that incorporate explicit robot descriptions [1, 6, 41] or infer embodiment properties from interaction history through in-context learning [32, 35]. Richer cross-embodiment datasets would enable systematic investigation of these designs.

How to Evaluate Transfer?

Scientific progress in cross-embodiment learning requires clearer evaluation of transfer. Calling a target robot “unseen” says little about the difficulty of generalization. Studies should specify how it differs from the training population, the control interface used, and the assumptions made about the task. They should also report what embodiment information and interaction data are available at test time, distinguishing zero-shot generalization from few-shot adaptation.

Task and embodiment are distinct but connected axes of generalization. Experience on a robot performing other tasks may already reveal its sensing, dynamics, and control responses. A policy can combine this experience with task information from other robots to solve a held-out task-embodiment combination [42]. Evaluation should distinguish this setting from transfer to an entirely unseen embodiment where no data is ever collected. Determining which “calibration experiences” on a new robot most improve performance across other tasks is itself an important research question.

Evaluation should also distinguish the behavioral changes required for transfer. Motion-level transfer preserves task-space motion or contact patterns by adapting joint-level execution. Strategy-level transfer requires a different way of achieving the same goal. For example, shaking may unfold a garment effectively with one embodiment but exceed another robot’s dynamic capabilities, making slower, sequential unfolding more appropriate. This requires a policy to relate the task’s intended outcome to the robot’s capabilities and select a feasible interaction strategy. Preserving a motion and discovering an alternative strategy therefore test different transfer capabilities.

Simulation offers a practical platform for these evaluations. Existing work demonstrates its value for scalable policy evaluation [33, 56], and dedicated benchmarks have begun to assess generalization across morphologies [40]. Procedural generation could extend these efforts to diverse, mechanically plausible embodiments with systematic variation in sensing, morphology, dynamics, actuation, and control. Controlled comparisons would help accurately characterize a policy’s transfer behaviors, isolating which differences a policy accommodates, where transfer fails, and how these limits change as training experience grows.

Towards Adaptive and Self-Evolving Robots

Cross-embodiment learning and transfer create broad possibilities for robotics. In nature, physical capabilities develop through adaptation within a lifetime and evolution across generations. Learning that accumulates across changing embodiments could connect these processes in artificial systems, allowing intelligence to increasingly shape the physical means through which it acts.

One route toward this vision is computational robot design. Computational co-design already optimizes physical form and control together, from simulated evolution [47] and automatically designed and fabricated machines [34] to joint optimization of morphology and policy [7, 8, 19, 44]. A central bottleneck is that evaluating each candidate embodiment requires establishing what it can achieve under an appropriate controller. Efficient cross-embodiment transfer could reduce this cost, enabling broader searches over morphology, sensing, and actuation for more demanding tasks. Designs could then be evaluated both for their immediate performance and for how readily they support learning new capabilities.

Computational robot design could become a part of the co-evolution process of embodiment and policy. Better policies could make previously difficult de-

signs viable, and better designs could make new behaviors easier to learn. Computational studies inspired by biological evolution already show that evolved morphologies can facilitate learning and adaptation to new tasks [18]. Cross-embodiment transfer could carry learned capabilities across successive designs, allowing each generation to build on experience acquired by earlier ones. New embodiments would, in turn, expose the learner to interactions and capabilities unavailable on previous platforms. Policy learning and embodiment design could thereby improve together, producing forms and behaviors better suited to the tasks and environments we care about.

Within an agent’s lifetime, embodiment could also become a means of adaptation. Biology offers a concrete precedent in red knots, which reversibly change their gizzard size in response to changes in diet [10]. Artificial agents could adjust their physical configurations to changing demands by selecting tools, repositioning sensors, or reconfiguring limbs to gain reach, stability, or access. Cross-embodiment control could help them exploit these changes without learning each configuration from scratch. Selecting and modifying an embodiment would then become part of how an agent pursues its goals.

The practical challenge of reusing experience across robots thus leads to a broader scientific pursuit. Cross-embodiment learning could provide a foundation for physical intelligence that persists as its embodiments change and gains agency over those changes. Nature offers a precedent for the diversity of capabilities that adaptation and evolution can produce. The ambition of intelligence across embodiments is to build artificial agents that continually expand both what they can physically achieve and the physical means through which they can accomplish goals.

Acknowledgements

We thank Zhanxin Wu, Marcel Torne, and Nico Bohlinger for helpful feedback.

References

1. Ai, B., Dai, L., Bohlinger, N., Li, D., Mu, T., Wu, Z., et al.: Towards Embodiment Scaling Laws in Robot Locomotion. In: Conference on Robot Learning (CoRL) (2025)
2. Ai, B., Tian, S., Shi, H., Wang, Y., Pfaff, T., Tan, C., et al.: A Review of Learning-Based Dynamics Models for Robotic Manipulation. *Science Robotics* **10**(106), eadt1497 (2025). DOI 10.1126/scirobotics.adt1497
3. Amato, N.M., Hutchinson, S., Garg, A., Billard, A., Rus, D., Tedrake, R., et al.: “Data will solve robotics and automation: True or false?": A debate. *Science Robotics* **10**(105), eaea7897 (2025). DOI 10.1126/scirobotics.aea7897. URL <https://www.science.org/doi/10.1126/scirobotics.aea7897>
4. Anthropic: Claude Fable 5.1. Anthropic model release (2026). URL <https://www.anthropic.com/claude/fable>

5. Barreiros, J., Beaulieu, A., Bhat, A., Cory, R., Cousineau, E., Dai, H., et al.: A careful examination of large behavior models for multitask dexterous manipulation. *Science Robotics* **11**(113), eaea6201 (2026). DOI 10.1126/scirobotics.aea6201. URL <https://www.science.org/doi/abs/10.1126/scirobotics.aea6201>
6. Bohlinger, N., Czechmanowski, G., Krupka, M., Kicki, P., Walas, K., Peters, J., et al.: One Policy to Run Them All: an End-to-end Learning Approach to Multi-Embodiment Locomotion. *Conference on Robot Learning* (2024)
7. Bohlinger, N., Peters, J.: Shape your body: Value gradients for multi-embodiment robot design. *arXiv preprint arXiv:2606.00702* (2026)
8. Chen, T., He, Z., Ciocarlie, M.: Hardware as policy: Mechanical and computational co-optimization using deep reinforcement learning. In: *Proceedings of the 2020 Conference on Robot Learning, Proceedings of Machine Learning Research*, vol. 155, pp. 1158–1173 (2021). URL <https://proceedings.mlr.press/v155/chen21a.html>
9. Chi, C., Xu, Z., Pan, C., Cousineau, E., Burchfiel, B., Feng, S., et al.: Universal Manipulation Interface: In-The-Wild Robot Teaching Without In-The-Wild Robots. In: D. Kulic, G. Venture, K.E. Bekris, E. Coronado (eds.) *Robotics: Science and Systems XX*, Delft, The Netherlands, July 15-19, 2024 (2024). DOI 10.15607/RSS.2024.XX.045. URL <https://doi.org/10.15607/RSS.2024.XX.045>
10. Dekinga, A., Dietz, M.W., Koolhaas, A., Piersma, T.: Time course and reversibility of changes in the gizzards of red knots alternately eating hard and soft food. *Journal of Experimental Biology* **204**, 2167–2173 (2001). DOI 10.1242/jeb.204.12.2167
11. Deshpande, A., Guru, M., Hendrix, R., Jauhri, S., Eftekhari, A., Tripathi, R., et al.: MolmoB0T: Large-Scale Simulation Enables Zero-Shot Manipulation (2026). URL <https://arxiv.org/abs/2603.16861>
12. Dyna Robotics: Dyna-2: A 1-million-hour scaling law for world-action models (2026). URL <https://dyna.co/dyna-2>
13. Fang, H.S., Romero, B., Xie, Y., Hu, A., Huang, B.R., Alvarez, J., et al.: DEXOP: A device for robotic transfer of dexterous human manipulation. *arXiv preprint arXiv:2509.04441* (2025)
14. Fu, Z., Zhao, Q., Wu, Q., Wetzstein, G., Finn, C.: HumanPlus: Humanoid shadowing and imitation from humans. In: *Proceedings of the 8th Conference on Robot Learning, Proceedings of Machine Learning Research*, vol. 270, pp. 2828–2844 (2025)
15. Fu, Z., Zhao, T.Z., Finn, C.: Mobile ALOHA: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. *arXiv preprint arXiv:2401.02117* (2024)
16. Generalist Team: Towards Machines with a Thousand Hands. Generalist AI research blog (2026). URL <https://generalistai.com/blog/towards-machines-with-a-thousand-hands>
17. Goldberg, K.: Good old-fashioned engineering can close the 100,000-year “data gap” in robotics. *Science Robotics* **10**(105), eaea7390 (2025). DOI 10.1126/scirobotics.aea7390. URL <https://www.science.org/doi/abs/10.1126/scirobotics.aea7390>
18. Gupta, A., Savarese, S., Ganguli, S., Fei-Fei, L.: Embodied intelligence via learning and evolution. *Nature Communications* **12**(1), 5721 (2021). DOI 10.1038/s41467-021-25874-z. URL <https://www.nature.com/articles/s41467-021-25874-z>
19. Ha, H., Liu, C.K., Song, S.: Transformer Transformer: A Unified Model for Motion-Conditioned Robot Co-design (2026). URL <https://arxiv.org/abs/2607.25798>
20. He, T., Luo, Z., He, X., Xiao, W., Zhang, C., Zhang, W., et al.: OmniH2O: Universal and dexterous human-to-humanoid whole-body teleoperation and learning.

- In: Proceedings of the 8th Conference on Robot Learning, *Proceedings of Machine Learning Research*, vol. 270, pp. 1516–1540 (2025)
21. He, Z., Ai, B., Mu, T., Liu, Y., Wan, W., Fu, J., et al.: Scaling Cross-Embodiment World Models for Dexterous Manipulation. *arXiv* (2025)
 22. Junge, K., Hughes, J.: ADAPT-Teleop: Robotic hand with human matched embodiment enables dexterous teleoperated manipulation. *npj Robotics* **3**(1), 31 (2025). DOI 10.1038/s44182-025-00034-3
 23. Kareer, S., Pertsch, K., Darpinian, J., Hoffman, J., Xu, D., Levine, S., et al.: Emergence of Human to Robot Transfer in Vision-Language-Action Models (2025). URL <https://arxiv.org/abs/2512.22414>
 24. Kaufmann, E., Bauersfeld, L., Loquercio, A., Müller, M., Koltun, V., Scaramuzza, D.: Champion-level drone racing using deep reinforcement learning. *Nature* **620**(7976), 982–987 (2023)
 25. Khazatsky, A., Pertsch, K., Nair, S., Balakrishna, A., Dasari, S., Karamcheti, S., et al.: DROID: A Large-Scale In-The-Wild Robot Manipulation Dataset. In: D. Kulić, G. Venture, K.E. Bekris, E. Coronado (eds.) *Robotics: Science and Systems XX*, Delft, The Netherlands, July 15–19, 2024 (2024). DOI 10.15607/RSS.2024.XX.120. URL <https://doi.org/10.15607/RSS.2024.XX.120>
 26. Kirschner, R.J., Karacan, K., Melone, A., Haddadin, S.: Categorizing robots by performance fitness into the tree of robots. *Nature Machine Intelligence* **7**(3), 459–470 (2025)
 27. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks. In: *Advances in Neural Information Processing Systems*, vol. 25 (2012). URL https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
 28. Kumar, A., Fu, Z., Pathak, D., Malik, J.: RMA: Rapid motor adaptation for legged robots. *arXiv preprint arXiv:2107.04034* (2021)
 29. LaValle, S.M.: *Planning Algorithms*. Cambridge University Press (2006). URL <https://lavalle.pl/planning/>
 30. Levine, S., Finn, C., Darrell, T., Abbeel, P.: End-to-End Training of Deep Visuomotor Policies. *Journal of Machine Learning Research* **17**(39), 1–40 (2016). URL <https://www.jmlr.org/papers/v17/15-522.html>
 31. Li, B., Yin, X., Lin, M., Zhang, Y., Xu, D.: Egowam: World action models beyond pixels with in-the-wild egocentric human data. *arXiv preprint arXiv:2607.08436* (2026)
 32. Li, D., Ai, B., Bohlinger, N., Peters, J., Su, H., Christensen, H.I.: Rapid embodiment adaptation for quadrupedal locomotion. *arXiv preprint arXiv:2608.01506* (2026)
 33. Li, X., Hsu, K., Gu, J., Mees, O., Pertsch, K., Walke, H.R., et al.: Evaluating real-world robot manipulation policies in simulation. In: Proceedings of The 8th Conference on Robot Learning, *Proceedings of Machine Learning Research*, vol. 270, pp. 3705–3728. PMLR (2025). URL <https://proceedings.mlr.press/v270/li25c.html>
 34. Lipson, H., Pollack, J.B.: Automatic Design and Manufacture of Robotic Lifeforms. *Nature* **406**(6799), 974–978 (2000). DOI 10.1038/35023115
 35. Liu, M., Pathak, D., Agarwal, A.: LocoFormer: Generalist Locomotion via Long-context Adaptation (2025). URL <https://arxiv.org/abs/2509.23745>
 36. Losos, J.B.: Ecomorphology, performance capability, and scaling of west indian anolis lizards: An evolutionary analysis. *Ecological Monographs* **60**(3), 369–388 (1990). DOI 10.2307/1943062

37. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2), 91–110 (2004). DOI 10.1023/B:VISI.0000029664.99615.94
38. Lynch, K.M., Park, F.C.: *Modern Robotics: Mechanics, Planning, and Control*. Cambridge University Press (2017)
39. OpenAI: GPT-6 Astra: A New Generation of Intelligence. OpenAI model release (2026). URL <https://openai.com/index/gpt-6-astra/>
40. Parakh, M., Kirchmeyer, A., Han, B., Deng, J.: AnyBody: A benchmark suite for cross-embodiment manipulation. *arXiv preprint arXiv:2505.14986* (2025). URL <https://arxiv.org/abs/2505.14986>
41. Patel, A., Song, S.: GET-Zero: Graph Embodiment Transformer for Zero-shot Embodiment Generalization. In: 2025 IEEE International Conference on Robotics and Automation (ICRA) (2025)
42. Physical Intelligence: $\pi_{0.7}$: a Steerable Generalist Robotic Foundation Model with Emergent Capabilities. *arXiv preprint arXiv:2604.15483* (2026)
43. Pisenò, M., Tevet, G., Liu, C.K.: Cloak: Zero-shot cross-embodiment manipulation by masking the end-effector from the vla. *arXiv preprint arXiv:2606.22836* (2026)
44. Schaff, C., Yunis, D., Chakrabarti, A., Walter, M.R.: Jointly learning to construct and control agents using deep reinforcement learning. In: 2019 international conference on robotics and automation (ICRA), pp. 9798–9805. IEEE (2019)
45. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., et al.: Mastering the game of Go with deep neural networks and tree search. *nature* **529**(7587), 484–489 (2016). DOI 10.1038/nature16961
46. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., et al.: A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science* **362**(6419), 1140–1144 (2018). DOI 10.1126/science.aar6404
47. Sims, K.: Evolving virtual creatures. In: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, pp. 15–22 (1994). DOI 10.1145/192161.192167
48. Sutton, R.S.: The Bitter Lesson. Essay, March 13 (2019). URL <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>
49. Thrun, S., Burgard, W., Fox, D.: *Probabilistic Robotics*. MIT Press (2005)
50. Tian, T., Li, H., Ai, B., Yuan, X., Huang, Z., Su, H.: Diffusion Dynamics Models with Generative State Estimation for Cloth Manipulation. In: Conference on Robot Learning (CoRL) (2025)
51. Wang, C., Shi, H., Wang, W., Zhang, R., Fei-Fei, L., Liu, C.K.: Dexcap: Scalable and portable mocap data collection system for dexterous manipulation. *arXiv preprint arXiv:2403.07788* (2024)
52. Wang, Q., Fang, K.: KITE: Decoupling Kinematics and Interaction for Zero-Shot Cross-Embodiment Manipulation (2026). URL <https://arxiv.org/abs/2606.22113>
53. Xu, M., Xu, Z., Xu, Y., Chi, C., Wetzstein, G., Veloso, M., et al.: Flow as the Cross-Domain Manipulation Interface. In: Conference on Robot Learning (2024)
54. Xu, M., Zhang, H., Hou, Y., Xu, Z., Fan, L., Veloso, M., et al.: DexUMI: Using Human Hand as the Universal Manipulation Interface for Dexterous Manipulation. In: Proceedings of The 9th Conference on Robot Learning, *Proceedings of Machine Learning Research*, vol. 305, pp. 437–459. PMLR (2025). URL <https://proceedings.mlr.press/v305/xu25b.html>
55. Zha, L., Hancock, A.J., Zhang, M., Yin, T., Huang, Y., Shah, D., et al.: LAP: Language-action pre-training enables zero-shot cross-embodiment transfer. *arXiv preprint arXiv:2602.10556* (2026)

- 56. Zhang, K., Sha, S., Jiang, H., Loper, M., Song, H., Cai, G., et al.: Real-to-sim robot policy evaluation with gaussian splatting simulation of soft-body interactions. arXiv preprint arXiv:2511.04665 (2025). URL <https://arxiv.org/abs/2511.04665>
- 57. Zhao, A., Xu, J., Konaković-Luković, M., Hughes, J., Spielberg, A., Rus, D., Matusik, W.: RoboGrammar: Graph Grammar for Terrain-Optimized Robot Design. *ACM Transactions on Graphics* **39**(6), 188:1–188:16 (2020). DOI 10.1145/3414685.3417831. URL <https://doi.org/10.1145/3414685.3417831>
- 58. Zhao, T.Z., Kumar, V., Levine, S., Finn, C.: Learning fine-grained bimanual manipulation with low-cost hardware. *Robotics: Science and Systems* (2023)