# TidyBot++: An Open-Source Holonomic Mobile Manipulator for Robot Learning

**Anonymous Author(s)**
Affiliation
Address
`email`

**Abstract:** Exploiting the promise of recent advances in imitation learning for mobile manipulation will require the collection of large numbers of human-guided demonstrations. This paper proposes an open-source design for an inexpensive, robust, and flexible mobile manipulator that can support arbitrary arms, enabling a wide range of real-world household mobile manipulation tasks. Crucially, our design uses powered casters to enable the mobile base to be fully *holonomic*, able to control all planar degrees of freedom independently and simultaneously. This feature makes the base more maneuverable and simplifies many mobile manipulation tasks, eliminating the kinematic constraints that create complex and time-consuming motions in nonholonomic bases. We equip our robot with an intuitive mobile phone teleoperation interface to enable easy data acquisition for imitation learning. In our experiments, we use this interface to collect data and show that the resulting learned policies can successfully perform a variety of common household mobile manipulation tasks.

**Keywords:** mobile manipulation, imitation learning, holonomic drive

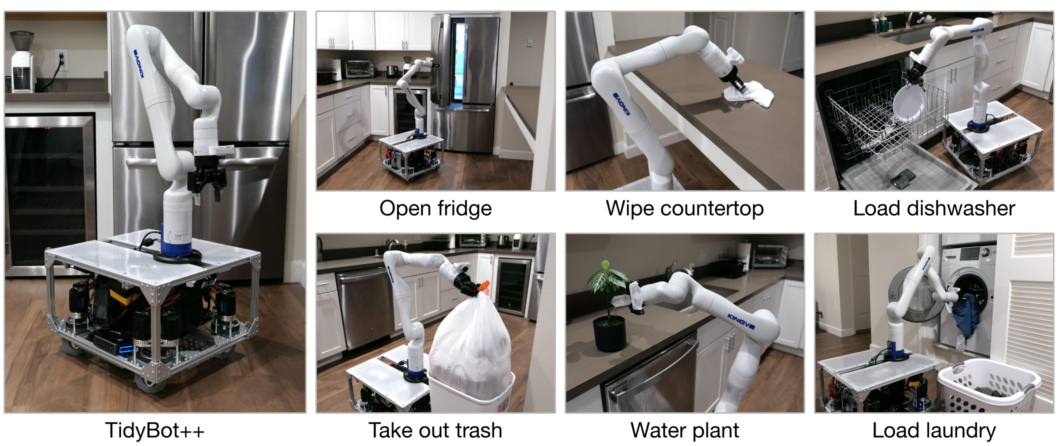


Figure 1: We develop an open-source mobile manipulator with a holonomic base (left), and show that it can perform a variety of household tasks in a real apartment home (right).

## 1 Introduction

Imitation learning from real-world data is starting to show very promising results in robotics for both fixed-arm robots [1, 2, 3, 4, 5, 6, 7] and mobile manipulators [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. However, one key bottleneck is the availability of data. Unlike in natural language processing, which can train on readily-available data from the internet, real-world data for training

Submitted to the 8th Conference on Robot Learning (CoRL 2024). Do not distribute.

robot policies is much harder to come by. As a result, scaling data collection of robotic tasks has become a high-interest research direction. Recent efforts have collected large robot learning datasets to address this [22, 23, 24, 25, 26, 27]. These datasets were largely collected on fixed-arm robot setups. However, to bring robots to their full potential, mobility is important as it will enable many more tasks in realistic household settings [8].

We believe that one reason there are so few data collection and learning efforts in mobile manipulation is the lack of suitable research hardware. Existing commercial options for mobile bases are often tailored towards industrial or warehouse use cases, and may be ill-suited for household environments due to their large size. They are also often expensive and are typically subject to kinematic constraints.

In this work, we propose an open-source design for a mobile base designed to carry a robot arm sized for use in household environments. In addition to being inexpensive, flexible, and easy to assemble, our base is *holonomic*: able to independently move in any of the three degrees of freedom (DoFs) on the ground plane—$(x, y, \theta)$—at any time. We argue that this is an important advantage for more intuitive teleoperation, and greatly increases the ease of acquiring large amounts of training data for real-world imitation learning.

Nonholonomic robots, such as differential drive (wheelchair-like) or Ackermann drive (car-like) platforms, have constrained degrees of freedom in their motion. The most notable consequence of this is that they cannot move sideways. For example, cars cannot directly drive sideways into a street-side parking spot and must execute a multi-step parallel-parking maneuver.

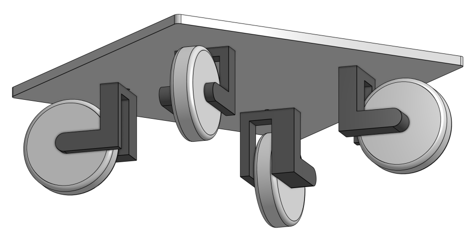

Figure 2: A simplified illustration of caster wheels on a holonomic base.

In contrast, holonomic robots have no kinematic constraints and can simultaneously and independently control all three degrees of freedom. An example of a holonomic vehicle common in everyday life is an office chair, which can be smoothly pushed or rotated in arbitrary directions. This is enabled by the design of the caster wheels (Fig. 2), which have an offset between the vertical axis of the swivel mechanism and the roll axis of the wheel. This offset is a crucial design feature of casters and is what makes the office chair fully holonomic. It creates a lever arm that causes the wheel to trail behind the vertical axis of the swivel as the chair moves, automatically aligning the wheels to the direction of movement. Without a caster offset, the vehicle would be omnidirectional (capable of moving in any direction) but still nonholonomic, as the wheels have to be manually aligned to face the direction of desired motion before the vehicle can begin moving. Overall, holonomic drive is preferred for maximum maneuverability.

Our holonomic base uses a powered-caster drive mechanism [28]. It is driven by four motorized casters, and can be thought of as a motorized office chair. The ability to steer all four wheels makes the base omnidirectional, and the caster offset makes the base holonomic, allowing it to instantaneously accelerate in any direction as it does not need to first align the wheels to the direction of motion.

A holonomic mobile base enables easier teleoperation and kinesthetic teaching for collecting imitation learning data. Everyday tasks such as opening doors and cabinets often require sideways motions of the mobile base to improve the workspace of the arm during execution. This useful motion is not immediately available with a differential drive base. Instead, the robot has to replan vehicle trajectories to satisfy nonholonomic constraints, which costs extra motion and time with no added value to the task. A holonomic mobile base, on the other hand, can be much more reactive. It can be moved arbitrarily in any direction no matter the current configuration, allowing an operator to make fine adjustments to the positioning of the base.

A holonomic base is also useful for policy learning and inference. Recent real-robot imitation learning works have converged on the use of position representations, as they are more stable and less noisy compared to velocities. However, a nonholonomic mobile base can only be controlled in velocity mode [8, 16]. A holonomic base, on the other hand, can be directly commanded to go to a task space position $(x, y, \theta)$ in a repeatable manner, as it can independently control all DOFs with no constraints. In our experiments, we show that we can indeed train high-performing policies for our robot across several mobile manipulation tasks in a real apartment home. Additionally, we show that policies can be learned more easily with data collected from a holonomic base compared to a nonholonomic one.

To facilitate easy data collection with our new mobile manipulator, we also develop a mobile phone teleoperation interface. The interface uses the WebXR API [29] to stream the real-time 6-DoF pose of the mobile phone to a computer, which maps the phone motion to corresponding motions of the mobile base or arm via low-level control. WebXR is supported on most modern Android and iOS phones, so our interface does not require purchase of a separate teleoperation device. In our experiments, we use this teleoperation interface to collect data for training our policies.

Our holonomic mobile base is low-cost ($5–6k USD) and designed from the ground up to optimize for mobile manipulation research productivity. We will fully open source all aspects of this system, including the hardware design, mobile phone teleoperation interface, policy learning setup, and low-level controller. We will also create a documentation webpage for the mobile base, including bill of materials (BOM), a hardware assembly guide with videos, and 3D CAD files. We believe these components can help democratize access to highly maneuverable mobile manipulators, increase ease and practicality of mobile manipulation data collection, and improve research reproducibility by providing a standardized and reusable robot platform.

Our key contributions in this work are thus: (1) an open-source design for a holonomic mobile manipulator, (2) a mobile phone teleoperation interface for easily collecting data with the mobile manipulator, and (3) demonstration that our system is capable of learning policies.

## 2 Related Work

**Data collection for mobile manipulation.** To address the lack of robotics data for learning manipulation policies, several works have developed data collection platforms. The majority of these platforms are built for fixed-arm setups [30, 31, 25, 32]. For example, the DROID dataset [25] was collected on a standardized setup with an arm mounted on a portable table. The authors use an Oculus controller to teleoperate the robot. However, this controller must remain in view of four IR receivers which can lead to unexpected motion if the controller moves out of view and back. Similar to our work, RoboTurk [30, 33] used a mobile phone to teleoperate fixed-base robot arms which is a much more flexible solution and does not require purchasing a dedicated teleoperation device. However, their system suffers from drift as they only rely on IMU measurements and do not use the camera. MART [32] and MOMART [34], which extend RoboTurk to multi-arm and mobile manipulation, respectively, suffer from similar shortcomings and have not been demonstrated on real robots. In our work, we use the WebXR API [29], which combines IMU data with visual odometry based on the phone's camera to mitigate drift. TeleMoMa [35] is a teleoperation framework that supports multiple teleoperation interfaces and three commercially-available, high-cost robots (for a detailed comparison to our low-cost base, see Tab. **??**). One of the supported teleoperation devices is a mobile phone app based on ARKit, similar to what we use. Our interface is based on WebXR, which leverages ARKit on iPhone but works on Android as well.

There are several works that propose data collection devices that are hand-held by the human demonstrator [31, 11] but in this case the demonstrator does not get direct feedback on whether a demonstration is kinematically feasible by the robot. Of those approaches, Dobb·E [11] proposes a low-cost reacher-grabber stick with a mounted iPhone to record data. The authors then train visuomotor policies on this data that are deployed on a differential drive Stretch robot [36]. Mobile ALOHA [8] is a dual-arm mobile manipulation platform capable of performing an impressive array of household tasks. However, the robot's differential drive base and large footprint limits its maneuverability, and

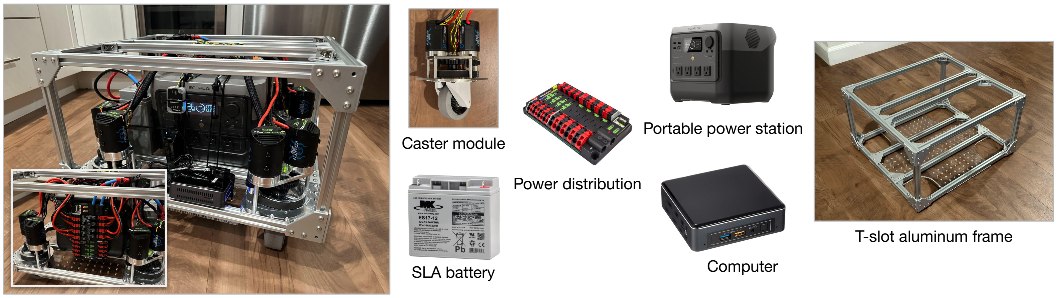


Figure 3: Our mobile base is designed to be modular and easily reconfigurable. It has very few components and can be assembled in 1 to 2 days.

the arms are not able to reach the ground. Furthermore, the teleoperator is strapped to the back of the platform far away from the end effectors, which can make it hard to teleoperate precise actions. For our system, the teleoperator can freely walk around the scene and get very close when precision is required.

## 3 Hardware Design

We designed this mobile base concept from the ground up to optimize for mobile manipulation research productivity. It is simple, low-cost ($5–6k), and modular (Fig. 3). The core is the drive system, which is based on readily available components from the FIRST Robotics Competition (FRC) [37] ecosystem. A basic frame built out of aluminum T-slot extrusions carries the four motorized caster modules that are powered through a fused power distribution panel by a sealed lead acid (SLA) battery. There are many similar components in the FRC ecosystem that could be used to build similar systems. The large and active community of FRC users and vendors ensures that components are well-documented and readily available.

## 4 Experiments

In these experiments, we aim to show that our teleoperation interface can collect useful demonstration data to successfully train policies for a variety of household mobile manipulation tasks.

### 4.1 Imitation learning

We used our phone teleoperation interface to collect demonstrations for the 6 tasks shown in Tab. 1. We collected 100 demonstrations for the shorter *open fridge* task and 50 for all others. Data collection for each task took between 1 and 2 hours for 50 episodes, including overhead for environment resets.

Table 1: Imitation learning results

| Task | Success rate |
|---|---|
| Open fridge | 10/10 |
| Wipe countertop | 9/10 |
| Load dishwasher | 7/10 |
| Take out trash | 10/10 |
| Load laundry | 7/10 |
| Water plant | 6/10 |

We then used the data to train a diffusion policy [1] for each task. We trained each policy for 500 epochs and evaluated them by running 10 episodes of policy rollouts. Success rates are shown in Tab. 1. Note that while diffusion policies are typically trained using 200 to 300 demonstrations, we found that 50 was already sufficient for the robot to learn to complete the task successfully. Performance can likely be further improved with more data. These results show that our system is capable of learning high-performing policies for useful tasks in real homes.

# References


[1] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[2] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.

[3] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel. Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators. *arXiv preprint arXiv:2309.13037*, 2023.

[4] C. Wang, L. Fan, J. Sun, R. Zhang, L. Fei-Fei, D. Xu, Y. Zhu, and A. Anandkumar. Mimicplay: Long-horizon imitation learning by watching human play. *arXiv preprint arXiv:2302.12422*, 2023.

[5] H. Ha, P. Florence, and S. Song. Scaling up and distilling down: Language-guided robot skill acquisition. In *Conference on Robot Learning*, pages 3766–3777. PMLR, 2023.

[6] J. W. Kim, T. Z. Zhao, S. Schmidgall, A. Deguet, M. Kobilarov, C. Finn, and A. Krieger. Surgical robot transformer (srt): Imitation learning for surgical tasks. *arXiv preprint arXiv:2407.12998*, 2024.

[7] A. Z. Ren, J. Lidard, L. L. Ankile, A. Simeonov, P. Agrawal, A. Majumdar, B. Burchfiel, H. Dai, and M. Simchowitz. Diffusion policy policy optimization. *arXiv preprint arXiv:2409.00588*, 2024.

[8] Z. Fu, T. Z. Zhao, and C. Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. In *arXiv*, 2024.

[9] A. Prasad, K. Lin, J. Wu, L. Zhou, and J. Bohg. Consistency policy: Accelerated visuomotor policies via consistency distillation. In *Robotics: Science and Systems*, 2024.

[10] S. Lee, Y. Wang, H. Etukuru, H. J. Kim, N. M. M. Shafiullah, and L. Pinto. Behavior generation with latent actions. *arXiv preprint arXiv:2403.03181*, 2024.

[11] N. M. M. Shafiullah, A. Rai, H. Etukuru, Y. Liu, I. Misra, S. Chintala, and L. Pinto. On bringing robots home. *arXiv preprint arXiv:2311.16098*, 2023.

[12] R. Yang, Y. Kim, A. Kembhavi, X. Wang, and K. Ehsani. Harmonic mobile manipulation. *arXiv preprint arXiv:2312.06639*, 2023.

[13] H. Etukuru, N. Naka, Z. Hu, S. Lee, J. Mehu, A. Edsinger, C. Paxton, S. Chintala, L. Pinto, and N. M. M. Shafiullah. Robot utility models: General policies for zero-shot deployment in new environments. *arXiv preprint arXiv:2409.05865*, 2024.

[14] A. Sridhar, D. Shah, C. Glossop, and S. Levine. Nomad: Goal masked diffusion policies for navigation and exploration. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 63–70. IEEE, 2024.

[15] K. Ehsani, T. Gupta, R. Hendrix, J. Salvador, L. Weihs, K.-H. Zeng, K. P. Singh, Y. Kim, W. Han, A. Herrasti, et al. Spoc: Imitating shortest paths in simulation enables effective navigation and manipulation in the real world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16238–16250, 2024.

[16] H. Xiong, R. Mendonca, K. Shaw, and D. Pathak. Adaptive mobile manipulation for articulated objects in the open world. *arXiv preprint arXiv:2401.14403*, 2024.

[17] J. Yang, C. Glossop, A. Bhorkar, D. Shah, Q. Vuong, C. Finn, D. Sadigh, and S. Levine. Pushing the limits of cross-embodiment learning for manipulation and navigation. *arXiv preprint arXiv:2402.19432*, 2024.

[18] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn. Humanplus: Humanoid shadowing and imitation from humans. *arXiv preprint arXiv:2406.10454*, 2024.

[19] X. Cheng, J. Li, S. Yang, G. Yang, and X. Wang. Open-television: Teleoperation with immersive active visual feedback. *arXiv preprint arXiv:2407.01512*, 2024.

[20] P. Sundaresan, Q. Vuong, J. Gu, P. Xu, T. Xiao, S. Kirmani, T. Yu, M. Stark, A. Jain, K. Hausman, et al. Rt-sketch: Goal-conditioned imitation learning from hand-drawn sketches. 2024.

[21] H. Ha, Y. Gao, Z. Fu, J. Tan, and S. Song. Umi on legs: Making manipulation policies mobile with manipulation-centric whole-body controllers. *arXiv preprint arXiv:2407.10353*, 2024.

[22] H. Walke, K. Black, A. Lee, M. J. Kim, M. Du, C. Zheng, T. Zhao, P. Hansen-Estruch, Q. Vuong, A. He, V. Myers, K. Fang, C. Finn, and S. Levine. Bridgedata v2: A dataset for robot learning at scale. In *Conference on Robot Learning (CoRL)*, 2023.

[23] H.-S. Fang, H. Fang, Z. Tang, J. Liu, J. Wang, H. Zhu, and C. Lu. Rh20t: A robotic dataset for learning diverse skills in one-shot. In *RSS 2023 Workshop on Learning for Task and Motion Planning*, 2023.

[24] O. X.-E. Collaboration, A. O'Neill, A. Rehman, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, A. Jain, A. Tung, A. Bewley, A. Herzog, A. Irpan, A. Khazatsky, A. Rai, A. Gupta, A. Wang, A. Kolobov, A. Singh, A. Garg, A. Kembhavi, A. Xie, A. Brohan, A. Raffin, A. Sharma, A. Yavary, A. Jain, A. Balakrishna, A. Wahid, B. Burgess-Limerick, B. Kim, B. Schölkopf, B. Wulfe, B. Ichter, C. Lu, C. Xu, C. Le, C. Finn, C. Wang, C. Xu, C. Chi, C. Huang, C. Chan, C. Agia, C. Pan, C. Fu, C. Devin, D. Xu, D. Morton, D. Driess, D. Chen, D. Pathak, D. Shah, D. Büchler, D. Jayaraman, D. Kalashnikov, D. Sadigh, E. Johns, E. Foster, F. Liu, F. Ceola, F. Xia, F. Zhao, F. V. Frujeri, F. Stulp, G. Zhou, G. S. Sukhatme, G. Salhotra, G. Yan, G. Feng, G. Schiavi, G. Berseth, G. Kahn, G. Yang, G. Wang, H. Su, H.-S. Fang, H. Shi, H. Bao, H. B. Amor, H. I. Christensen, H. Furuta, H. Walke, H. Fang, H. Ha, I. Mordatch, I. Radosavovic, I. Leal, J. Liang, J. Abou-Chakra, J. Kim, J. Drake, J. Peters, J. Schneider, J. Hsu, J. Bohg, J. Bingham, J. Wu, J. Gao, J. Hu, J. Wu, J. Wu, J. Sun, J. Luo, J. Gu, J. Tan, J. Oh, J. Wu, J. Lu, J. Yang, J. Malik, J. Silvério, J. Hejna, J. Booher, J. Tompson, J. Yang, J. Salvador, J. J. Lim, J. Han, K. Wang, K. Rao, K. Pertsch, K. Hausman, K. Go, K. Gopalakrishnan, K. Goldberg, K. Byrne, K. Oslund, K. Kawaharazuka, K. Black, K. Lin, K. Zhang, K. Ehsani, K. Lekkala, K. Ellis, K. Rana, K. Srinivasan, K. Fang, K. P. Singh, K.-H. Zeng, K. Hatch, K. Hsu, L. Itti, L. Y. Chen, L. Pinto, L. Fei-Fei, L. Tan, L. J. Fan, L. Ott, L. Lee, L. Weihs, M. Chen, M. Lepert, M. Memmel, M. Tomizuka, M. Itkina, M. G. Castro, M. Spero, M. Du, M. Ahn, M. C. Yip, M. Zhang, M. Ding, M. Heo, M. K. Srirama, M. Sharma, M. J. Kim, N. Kanazawa, N. Hansen, N. Heess, N. J. Joshi, N. Suenderhauf, N. Liu, N. D. Palo, N. M. M. Shafiullah, O. Mees, O. Kroemer, O. Bastani, P. R. Sanketi, P. T. Miller, P. Yin, P. Wohlhart, P. Xu, P. D. Fagan, P. Mitrano, P. Sermanet, P. Abbeel, P. Sundaresan, Q. Chen, Q. Vuong, R. Rafailov, R. Tian, R. Doshi, R. Mart'in-Mart'in, R. Baijal, R. Scalise, R. Hendrix, R. Lin, R. Qian, R. Zhang, R. Mendonca, R. Shah, R. Hoque, R. Julian, S. Bustamante, S. Kirmani, S. Levine, S. Lin, S. Moore, S. Bahl, S. Dass, S. Sonawani, S. Song, S. Xu, S. Haldar, S. Karamcheti, S. Adebola, S. Guist, S. Nasiriany, S. Schaal, S. Welker, S. Tian, S. Ramamoorthy, S. Dasari, S. Belkhale, S. Park, S. Nair, S. Mirchandani, T. Osa, T. Gupta, T. Harada, T. Matsushima, T. Xiao, T. Kollar, T. Yu, T. Ding, T. Davchev, T. Z. Zhao, T. Armstrong, T. Darrell, T. Chung, V. Jain, V. Vanhoucke, W. Zhan, W. Zhou, W. Burgard, X. Chen, X. Chen, X. Wang, X. Zhu, X. Geng, X. Liu, X. Liangwei, X. Li, Y. Pang, Y. Lu, Y. J. Ma, Y. Kim, Y. Chebotar, Y. Zhou, Y. Zhu, Y. Wu, Y. Xu, Y. Wang, Y. Bisk, Y. Dou, Y. Cho, Y. Lee, Y. Cui, Y. Cao, Y.-H. Wu, Y. Tang, Y. Zhu, Y. Zhang, Y. Jiang, Y. Li, Y. Li, Y. Iwasawa, Y. Matsuo, Z. Ma, Z. Xu, Z. J. Cui, Z. Zhang, Z. Fu, and Z. Lin. Open X-Embodiment: Robotic learning datasets and RT-X models. https://arxiv.org/abs/2310.08864, 2023.

[25] A. Khazatsky, K. Pertsch, S. Nair, A. Balakrishna, S. Dasari, S. Karamcheti, S. Nasiriany, M. K. Srirama, L. Y. Chen, K. Ellis, P. D. Fagan, J. Hejna, M. Itkina, M. Lepert, Y. J. Ma, P. T. Miller, J. Wu, S. Belkhale, S. Dass, H. Ha, A. Jain, A. Lee, Y. Lee, M. Memmel, S. Park, I. Radosavovic, K. Wang, A. Zhan, K. Black, C. Chi, K. B. Hatch, S. Lin, J. Lu, J. Mercat, A. Rehman, P. R. Sanketi, A. Sharma, C. Simpson, Q. Vuong, H. R. Walke, B. Wulfe, T. Xiao, J. H. Yang, A. Yavary, T. Z. Zhao, C. Agia, R. Baijal, M. G. Castro, D. Chen, Q. Chen, T. Chung, J. Drake, E. P. Foster, J. Gao, D. A. Herrera, M. Heo, K. Hsu, J. Hu, D. Jackson, C. Le, Y. Li, K. Lin, R. Lin, Z. Ma, A. Maddukuri, S. Mirchandani, D. Morton, T. Nguyen, A. O'Neill, R. Scalise, D. Seale, V. Son, S. Tian, E. Tran, A. E. Wang, Y. Wu, A. Xie, J. Yang, P. Yin, Y. Zhang, O. Bastani, G. Berseth, J. Bohg, K. Goldberg, A. Gupta, A. Gupta, D. Jayaraman, J. J. Lim, J. Malik, R. Martín-Martín, S. Ramamoorthy, D. Sadigh, S. Song, J. Wu, M. C. Yip, Y. Zhu, T. Kollar, S. Levine, and C. Finn. Droid: A large-scale in-the-wild robot manipulation dataset. *arXiv preprint arXiv:2403.12945*, 2024.

[26] O. M. Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, T. Kreiman, C. Xu, et al. Octo: An open-source generalist robot policy. *arXiv preprint arXiv:2405.12213*, 2024.

[27] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

[28] R. Holmberg and O. Khatib. Development and control of a holonomic mobile robot for mobile manipulation tasks. *The International Journal of Robotics Research*, 19(11):1066–1074, 2000.

[29] I. W. W. Group. Webxr device api, 2024. URL https://www.w3.org/TR/webxr/.

[30] A. Mandlekar, Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay, et al. Roboturk: A crowdsourcing platform for robotic skill learning through imitation. In *Conference on Robot Learning*, pages 879–893. PMLR, 2018.

[31] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song. Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. In *Proceedings of Robotics: Science and Systems (RSS)*, 2024.

[32] A. Tung, J. Wong, A. Mandlekar, R. Martín-Martín, Y. Zhu, L. Fei-Fei, and S. Savarese. Learning multi-arm manipulation through collaborative teleoperation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9212–9219, 2021. doi: 10.1109/ICRA48506.2021.9561491.

[33] A. Mandlekar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín. What matters in learning from offline human demonstrations for robot manipulation. *arXiv preprint arXiv:2108.03298*, 2021.

[34] J. Wong, A. Tung, A. Kurenkov, A. Mandlekar, L. Fei-Fei, S. Savarese, and R. Martín-Martín. Error-aware imitation learning from teleoperation data for mobile manipulation. In *Conference on Robot Learning*, pages 1367–1378. PMLR, 2022.

[35] S. Dass, W. Ai, Y. Jiang, S. Singh, J. Hu, R. Zhang, P. Stone, B. Abbatematteo, and R. Martin-Martin. Telemoma: A modular and versatile teleoperation system for mobile manipulation. *arXiv preprint arXiv:2403.07869*, 2024.

[36] H. Robot. Stretch open-source mobile manipualtor, 2024. URL https://hello-robot.com/stretch-3-product.

[37] FIRST (For Inspiration and Recognition of Science and Technology). FIRST robotics competition, 2024. URL https://www.firstinspires.org/robotics/frc.