Design Decisions That Matter: Modality, State, and Action Horizon in Imitation Learning

Brendan Chharawala[1,2], Joshua Li[1,2], Stephie Liu[1,2], Shawn Yang[1,2], Colin Bellinger[2], David Liu[2], Chang Shu[2], Yue Hu[1], Pengcheng Xi[2]

[1]University of Waterloo, Ontario, Canada
[2]National Research Council Canada, Ontario, Canada

Brendan.Chharawala@uwaterloo.ca, Pengcheng.Xi@nrc-cnrc.gc.ca

Abstract: Advances in generalist robot learning models have been fuelled by large-scale demonstration datasets, yet which data are most effective remains underexplored. In particular, the role of teleoperation modality in shaping demonstration quality and downstream learning performance is still poorly understood. In this work, we present a comparative study of two common teleoperation interfaces, VR controllers and a haptic device, for collecting robot demonstrations on a robot platform. We focus on two assistive manipulation tasks, surface wiping and lamp switching, and collect a dataset of 400 human demonstrations. To capture operator workload, each session is evaluated using the NASA Task Load Index. We fine-tune the Octo model on these datasets, systematically varying the inclusion of robot state information and action horizon length. Our results highlight clear differences in data quality across modalities and their downstream impact on imitation learning performance. This study contributes insights into what makes robot learning data "good" and provides guidance on data collection design for assistive manipulation.

Keywords: Assistive Robotics, Dataset Design, Robot Learning

1 Introduction

Recent advances in generalist robot learning have been driven by large-scale demonstration datasets collected across diverse robots, tasks, and teleoperation setups [1, 2, 3, 4, 5]. While these datasets have enabled impressive progress, fundamental questions remain: what kinds of data are most useful for training or fine-tuning these models, and how should such data be collected? One critical but underexplored factor is the choice of teleoperation modality, which can strongly influence demonstration quality through differences in ease of use, precision, feedback, and operator workload.

At the same time, current systems often remain brittle when deployed in real-world settings. They can struggle with distribution shifts, when data collected on one embodiment must be transferred to another, or when policies trained in simulation are deployed on physical robots. These limitations are especially evident in domains such as assistive robotics, where robustness and reliability are essential. Understanding how data properties, and in particular teleoperation modality, shape downstream policy learning is therefore a central question for building more effective and generalizable robot learning systems.

In this work, we take a data-centric perspective and study how the choice of teleoperation modality affects downstream imitation learning performance. Specifically, we collect demonstrations for two representative assistive tasks: wiping a table surface and turning a desk lamp on and off, using a UR10e robotic arm. To generate these demonstrations, we focus on two widely used teleoperation interfaces: 1) VR controllers (Meta Quest 3), which offer flexible spatial mapping and accessibility,

and 2) Haptic devices (Haply Inverse 3), which provide grounded force-feedback and open the door to integrating tactile sensing in the future. Our dataset includes 100 episodes per modality per task (400 episodes in total), complemented by subjective workload ratings using the NASA Task Load Index (NASA-TLX) [6].

On the model side, we fine-tune Octo [3], following [7] which established it as a representative generalist robotic policy. Octo balances efficiency and capability: it is expressive for generalist manipulation policies while remaining effective for fine-tuning on moderate-scale datasets. We use Octo's transformer-based architecture and modular design, which is explicitly designed to adapt to new embodiments and setups, making it well suited for studying data-centric design choices. In contrast, more recent models such as OpenVLA [8] or policies developed by Physical Intelligence [9] are significantly larger, require substantially more data, and present higher computational costs for fine-tuning, factors that make them less appropriate for our setting. Using Octo therefore allows us to systematically vary two key fine-tuning design factors, whether robot state inputs are included and the action horizon length, while keeping the experiments feasible and interpretable. This setup enables us to analyze how teleoperation modality, dataset quality, and fine-tuning strategies interact to shape downstream policy performance.

Our contributions are threefold:

- We present a comparative dataset of assistive task demonstrations collected via VR controllers and haptic devices, paired with NASA-TLX subjective workload measures.
- We analyze how teleoperation modality influences demonstration quality and model fine-tuning performance on the Octo policy, providing evidence for which modalities produce "better" robot learning data.
- We investigate data-related fine-tuning design choices (inclusion of robot states and action horizon length) to highlight the role of dataset properties in enabling real-world robot performance.

2 Related Work

Teleoperation for assistive data collection. In the context of assistive robotics, most prior work has focused on simulation frameworks or small-scale real-world systems, but relatively few have studied how teleoperation can be leveraged to collect demonstrations that directly support learning. For example, the HARMONIC dataset captured human participants teleoperating a robotic arm to perform an assistive eating task, while recording multimodal data such as eye gaze, EMG, stereo video, and robot control signals [10]. This dataset has been valuable for analyzing human–robot interaction and modelling user intent, but has not been directly benchmarked for policy learning. Similarly, recent work on robotic assisted feeding introduced a visual imitation network with an attention mechanism trained on teleoperated demonstrations, showing how teleoperation data can enable real-robot deployment in a constrained task [11].

Beyond these examples, large-scale datasets such as OXE [5] and DROID [4] have enabled generalist robot policies, but they do not isolate the role of teleoperation modality in assistive scenarios. To our knowledge, there has been little systematic investigation of how different teleoperation interfaces, such as VR controllers or haptic devices, affect the quality of demonstrations and the performance of downstream policies in assistive tasks. This gap motivates our study: by comparing VR and haptic teleoperation on real assistive manipulation tasks, we provide new evidence on how modality influences data quality and learning outcomes.

Fine-tuning strategies. With good data, effective adaptation strategies are essential for generalist manipulation policies. Octo [3] is a representative open generalist policy trained on OXE (˜800k episodes), designed for rapid adaptation to new observation and action spaces, making it a natural backbone for data-centric studies. A recent paper [7] systematically ablates key fine-tuning design choices (action space, policy head, supervision targets, and parameter subsets) and offers guidance for adapting Octo-style generalist manipulation policies (GMPs). Complementary lines of work [12]

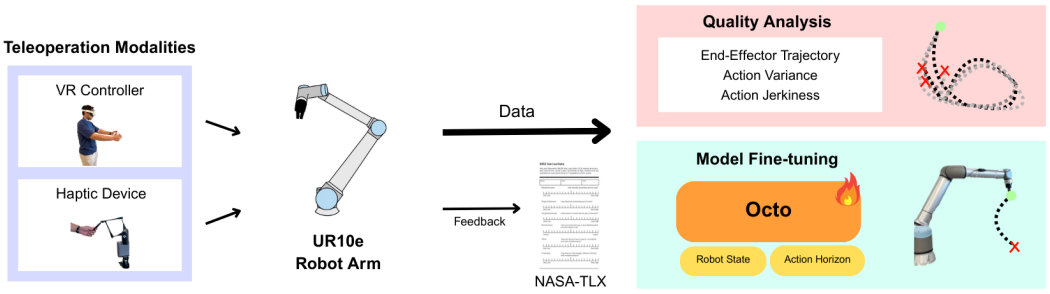

Figure 1: Overview of the data collection and learning pipeline. Teleoperation data are gathered on the UR10e robot arm using a VR controller and a haptic device, with user feedback collected via NASA-TLX. We then analyze the data using multiple quality metrics. To validate real-world performance, we further examine fine-tuning on Octo with each dataset individually and in combination, while systematically varying key design choices.

examine what data to fine-tune on. It directly compares demonstration modalities (kinesthetic, VR, spacemouse) and their downstream impact. Beyond Octo, approaches such as RoboFuME [13] and Diffusion Policy [14] motivate our own exploration of action horizon and state input choices during fine-tuning. In contrast to proposing new learners, our study quantifies how teleoperation modality and fine-tuning design factors interact to shape Octo's performance.

3 Methodology

Our methodology, illustrated in Fig. 1, is a structured pipeline designed to evaluate teleoperation modalities and design choices for robot policy learning. We collect demonstrations on a UR10e robot for two tasks: wiping a table surface and toggling a desk lamp. To promote data diversity and encourage model generalisation, the lamp is placed at five evenly spaced positions during data collection. These tasks are chosen as representative of assistive robotics applications: the wiping task emphasizes motions requiring a large range of movement, while the lamp task emphasizes precise manipulation (such as accurately engaging the pull chain without applying excessive force).

For each task, we collected demonstrations using two distinct teleoperation methods: a VR controller and a haptics-based 3D input device. Five participants were recruited (with Ethical Clearance in place), each contributing 20 episodes per modality per task, resulting in a total of 400 episodes across all modalities and tasks. To balance experience levels, participants were rotated across experiments and given a five-minute familiarisation period before recording. The participant pool included both male and female operators, with ages ranging from 21 to 63 years, and represented a mix of technical backgrounds and prior experience with robot controllers. This design ensured a consistent distribution of beginner, intermediate, and expert demonstrations for both modalities.

In addition, we ask users to complete NASA-TLX forms to measure subjective metrics that affect teleoperation usability. The demonstrations are converted into the standardized Reinforcement Learning Datasets (RLDS) format [15] to enable consistent downstream processing. We then perform data quality analysis across modalities to measure smoothness and control precision, focusing on metrics such as end-effector trajectories, action variance, and jerkiness. Finally, the collected data is used to fine-tune Octo, a state-of-the-art generalist robot policy, allowing us to assess how different input modalities influence downstream imitation learning performance.

3.1 Design Decisions

In imitation learning, the success of a policy depends not only on the learning algorithm but also on the *design of the dataset*. This involves two levels of decisions. At the collection level, the choice of teleoperation modality (e.g., VR vs. haptics) determines the fidelity, variability, and workload associated with the demonstrations [12]. At the usage level, representation decisions define how

demonstrations are consumed: whether to include detailed proprioceptive state information or more compact encodings, and how much temporal supervision to impose through the choice of action horizon (how many steps into the future, the policy predicts). Each of these factors changes the effective data distribution which the policy learns from.

While existing surveys comprehensively map methods, environments, and evaluation metrics in imitation learning [16], they provide limited guidance on how dataset design choices interact with operator workload and generalisation. We therefore pose the problem of identifying which dataset decisions (spanning modality, state inputs, and temporal structure) produce demonstrations that are both practical to collect and effective for training robust, consistently successful policies.

3.2 Teleoperation Modalities

VR Controller. Prior work [4] has already implemented the Oculus (VR) controller in data collection, and it has become a predominant methodology for robot teleoperation [5]. Our approach and setup are similar to [4]; we utilize a Meta Quest 3 headset and controller to read continuous positions. The operator movements are replicated on the robot such that the controller observations are mapped to the end effector pose. Gripper actions and robot movement status (enabled/disabled) are regulated by the controller buttons. Overall, the Oculus controller methodology allows for a large range of motion, and is ideal for completing broad tasks with great efficiency.

Haptic Device. The haptic data collection methodology utilizes a Haply Inverse3, which consists of a haptic system body, arms, and a pen (VerseGrip). Similar to the functionality of the VR controller, the operator can manipulate the pen's position to control the robot's end-effector pose. The pen buttons are associated with toggling the gripper action and robot movement status (enabled/disabled). Notably, the haptic device allows for finer controls due to its steadiness and responsiveness.

3.3 Subjective Metrics

After completing their demonstration episodes, each participant filled out a NASA-TLX survey. They rated six dimensions of workload, mental demand, physical demand, temporal demand, performance, effort, and frustration, on a 21-point Likert scale (0 = lowest workload, 20 = highest). Lower scores therefore indicate lower perceived workload.

3.4 Data Quality Metrics

Inspired by Li et al. [12], we employ a combination of qualitative and quantitative metrics to evaluate the quality of the demonstrated actions. Specifically, we utilize the end-effector trajectories, visualized in Cartesian coordinates (x, y, z), as our primary qualitative measure. For quantitative analysis, we adopt action variance and action jerkiness. Action variance is computed by comparing each action with the mean of actions from its K nearest state neighbours:

$$\text{ActionVariance}(D) = \frac{1}{|D|} \sum_{(s,a)\in D} (a - \frac{1}{K} \sum_{(\hat{s},\hat{a})\in NN(s,D,K)} \hat{a})^2 \tag{1}$$

Additionally, we calculate jerkiness as the second-order derivative of the action sequence.

3.5 Octo Robot Generalist Policy

We fine-tuned Octo Base 1.5 for 40k steps using demonstrations collected from both VR and haptic teleoperation. Training was performed with a cosine learning rate schedule peaking at 3×10^{-4}, a warmup of 2000 steps, weight decay of 0.01, and gradient clipping set to 1.0. We adopted a *language-conditioned full fine-tuning* setup, without freezing any model components. The input observations consisted of an isometric third-person view of the robot, a wrist-mounted camera view, and a short action history (current and previous step, window size = 2).

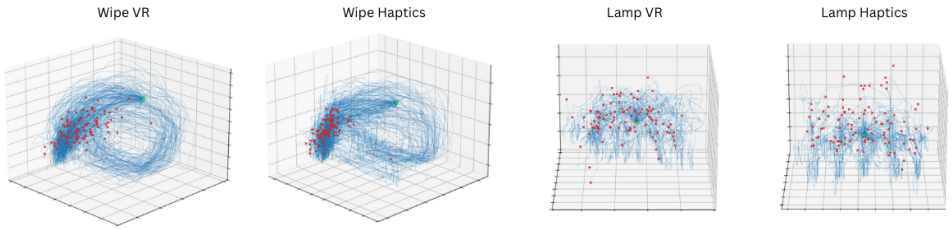

Figure 2: End-effector trajectories. VR produces more consistent trajectories for the wiping task, whereas the haptic device yields clearer pulling motions for the lamp task (five distinct vertical lines for five lamp positions).

Table 1: Action variance.

	VR	Haptics
Wipe	0.00057	0.00016
Lamp	0.00015	0.00006

Table 2: Jerkiness (mean $\pm$ std).

	VR	Haptics
Wipe	0.23 ± 0.33	0.12 ± 0.14
Lamp	0.12 ± 0.17	0.08 ± 0.08

4 Results and Analysis

4.1 Data Quality Results

To examine demonstration characteristics across teleoperation modalities, we visualize end-effector trajectories in Fig. 2. Visual inspection suggests that VR produces more consistent trajectories for the wiping task, whereas the haptic device yields clearer pulling motions for the lamp task, reflecting task requirements: wiping involves a large range of motion, while the lamp task requires precise manipulation.

As shown in Table 1 and 2, VR exhibits higher action variance and jerkiness across both tasks. Although this is expected, inferring controller positions via VR goggles is less stable than using a fixed haptic device, thus these quantitative differences do not fully capture data quality. In particular, differences in jerkiness are minimal, and the higher variance observed for VR on the wiping task can be largely due to the range of gripper ending positions (which occurs after task completion and is therefore inconsequential). Overall, VR provides higher-quality data for the wiping task, while haptic control is preferable for the lamp task.

4.2 NASA Task Load Index Results

Table 3 and 4 highlight clear differences across both tasks and modalities. Wiping was consistently more taxing than lamp switching, with higher mental demand (10.2 vs. 8.7), higher effort (9.6 vs. 8.7), and worse perceived performance (10.0 vs. 7.4). This matches intuition: wiping requires continuous, contact-rich motion, whereas lamp switching is short and discrete.

On comparing modalities, VR reduced operator burden, lowering physical demand (6.6 vs. 10.0) and temporal demand (7.7 vs. 9.0), but at the cost of lower perceived performance (9.6 vs. 7.8) and higher frustration (7.6 vs. 6.2). In general, VR was easier to use, but operators felt less precise and more irritated. Haptic control, though effortful, gave operators a greater sense of success and control.

These subjective experiences align with our earlier quantitative findings on action variance and jerkiness. VR produced noisier trajectories, which reflects the higher frustration ratings. Holistically, the results suggest a tradeoff: VR supports scalable data collection, while haptics yields higher-fidelity demonstrations applicable for training robust policies.

Mean pose error (cm) is computed at the onset of grasp or actuation as the Euclidean distance between the end-effector grasp frame and a task-specific target (the towel pinch point or the lamp

Table 3: NASA-TLX results per task (lower is better)

Task	Mental Dem.	Physical Dem.	Temporal Dem.	Performance	Effort	Frustr.
Wipe	10.2	8.4	8.6	10.0	9.6	6.3
Lamp	8.7	8.2	8.1	7.4	8.7	7.5

Table 4: NASA-TLX results per modality (lower is better)

Col. Method	Mental Dem.	Physical Dem.	Temporal Dem.	Performance	Effort	Frustr.
Haptics	10.0	10.0	9.0	7.8	9.2	6.2
VR	8.9	6.6	7.7	9.6	9.1	7.6

pull-switch), averaged across trials. Pose error serves as an indirect indicator of task failure, since deviations from the ideal location prevent correct execution (e.g., closing the gripper on the towel or pulling the lamp switch). These measurements complement success rates by providing insight into how lower-performing policies fail.

4.3 Success Rate and Pose Error on Model Deployment

Table 5: Success rate (%) and pose alignment error (cm). AH: action horizon; P: proprioception included. † Indicates low error due to oscillatory behaviour rather than task completion.

Configuration	Wipe		Lamp	
	Success (%)	Pose Err (cm)	Success (%)	Pose Err (cm)
Mixed, AH 10	**73**	**3.4**	**80**	**2.0**
VR, AH 10	47	4.7	53	3.4
Haptic, AH 10	40	4.7	53	3.2
Mixed, AH 15	27	4.6	47	4.6
Mixed, AH 5	20	10.5	33	5.3
Mixed, AH 10, P	20	9.7	13	3.7†

Table 5 summarises success rates across both tasks. Mixed-modality training (i.e., 100 episodes per task sampled randomly from 50% of the VR and haptic datasets, respectively) yields the highest performance, achieving 80% on the Lamp task and 73% on Wipe. These results substantially outperform single-modality training with VR (53% and 47%) or haptics (40% on Wipe). We also find that action horizons that are too short or too long (AH 5 or AH 15) reduce wiping success to 20% and 27%, respectively, underscoring horizon length as a key parameter influencing policy robustness. The strongest configurations rely on image observations combined with short action histories, while omitting explicit proprioceptive inputs, consistent with prior findings that excluding detailed proprioception can improve generalisation.

For the "Mixed, Action Horizon 10, Proprioception" configuration, the trained policy exhibited oscillatory behaviour, repeatedly moving around the lamp without converging. Although this behaviour produced a deceptively low pose error (since the end-effector frequently passed through the correct target location), the robot never stabilized long enough to complete the task successfully.

We evaluate the best-performing policies trained on each task and present their inference results in Fig. 3. On the left, the image sequence (top-to-bottom, left-to-right) illustrates a successful execution of the wiping task. On the right, the sequence (left-to-right) depicts a successful execution of the lamp toggling task.

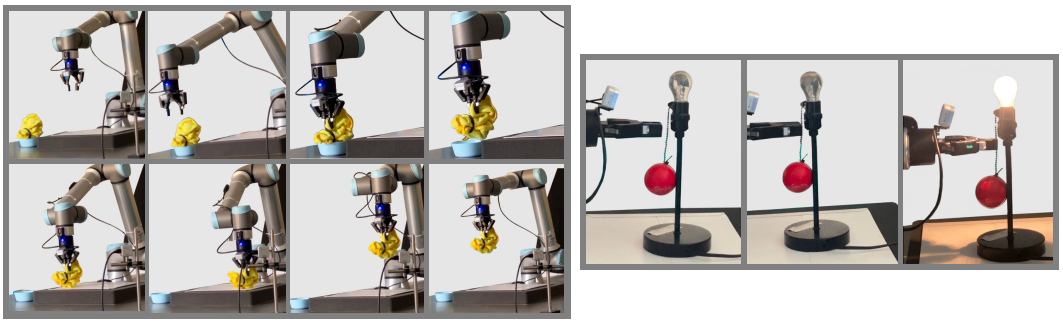

Figure 3: Inference results on the two tasks. The left panel shows a successful wiping episode, where the robot grasps a towel and wipes the table surface. The right panel shows the lamp task, where the robot grasps a thin string and pulls it to toggle the light.

5 Discussions

We treated data quality, sensor rig design, and training/evaluation alignment as first-order factors. We systematically varied two factors and assessed their impact on real-world robot behaviour: (i) inclusion of robot state inputs (joint angles, Tool Centre Point (TCP) pose, gripper status) and (ii) the action horizon length. Action horizon length strongly affected deployed performance on our 5.5 Hz control loop: horizons that were too short or too long degraded stability, while tuned values around 10 produced more reliable and consistent policies. Although only the first action is executed at inference, training with longer horizons provided richer temporal supervision that improved robustness. It should be recognized that offline validation metrics were insensitive to horizon length, underscoring that real-robot evaluation is essential for selecting horizon settings.

Surprisingly, policies trained without robot state inputs were more robust in our setup, improving approach accuracy and grasp initiation. We verified reasonable timing and values in the state streams, suggesting the effect likely stems from distribution or representation mismatches between the provided proprioceptive channels and the model's pretraining and camera configuration. This motivates selective inclusion or ablation of proprioceptive fields rather than assuming that "more state" always yields better generalisation.

Beyond these two controlled variations, several data- and rig-centric issues proved material. Fixing subtle pipeline bugs, most notably an RGB/BGR image-space mismatch between training and inference, and stale RNG keys reused across inference steps, produced clear stability gains. Camera placement and viewpoint also mattered: a rigid, third-person camera with broad tabletop coverage reduced self-occlusion, while the wrist camera exhibited visual aliasing near shallow approach angles, where qualitatively similar images corresponded to distinct end-effector poses. Finally, lighting variations in the Lamp task measurably affected policy stability despite standard photometric augmentations (unchanged from Octo Base), indicating that exposure control and task-specific augmentation remain important components of a data-first recipe for reliable deployment

Where we complement previous work [7] on strategies for training Octo: We corroborate several findings. Replacing the diffusion head with simpler linear heads yielded clear gains for us as well, supporting the claim that simpler heads can be more robust in practice. We also found, consistent with that work, that the "20-100 demos" guidance sits at the upper end in real deployments. Policies improved most as we approached the higher-count regime for number of demos.

Their conclusion in [12] that mixing teleoperation modalities can improve robustness aligns with our experience. We used VR and a 3D haptic controller and saw indications that heterogeneous demonstrations diversify state-action coverage and mitigate policy myopia. Our horizon-control experiments further suggest that temporal consistency mechanisms (beyond data diversification) are important for stable gripper actuation. This is an aspect that complements their data-collection focus with an inference-time control insight. New observations that refine the literature include:

(1) horizon length "sweet spots" matter when execution frequency is low, (2) selective proprio can outperform full proprio, indicating that "more sensors" is not always better, and (3) lighting remains a significant residual domain gap even with standard augmentations, which motivates more targeted photometric randomization and/or exposure control.

6 Conclusion

This work examined how data and design choices shape imitation learning performance for assistive manipulation. Using 400 demonstrations collected with VR and haptic teleoperation across wiping and lamp-switching tasks, paired with NASA-TLX ratings, we fine-tuned a generalist policy (Octo) while varying robot state inputs and action horizon. We found that these seemingly simple choices matter substantially in deployment: mixed-modality datasets with a tuned action horizon achieved the highest real-robot success (up to 73% on wiping and 80% on lamp), while offline metrics were largely insensitive to horizon length, underscoring the importance of real-robot evaluation. Modality trade-offs were consistent across quantitative and subjective measures: VR facilitated scalable collection with lower physical/temporal demand but noisier trajectories and higher frustration, whereas haptics yielded higher-fidelity demonstrations and better perceived performance.

Our design takeaways are data-centric. First, action horizon is a critical factor with task- and loop-frequency-dependent "sweet spots" (around 10 steps at $\sim$5.5 Hz here); too short or too long degraded stability. Second, selectively excluding detailed proprioceptive inputs improved robustness in our setup, suggesting that more sensors are not always better if misaligned with pretraining or the camera configuration. Third, rig and environment factors, camera viewpoint, near-contact visual aliasing from the wrist camera, and lighting variability, materially affected reliability despite standard augmentations. Together, these findings argue for a practical recipe for assistive settings: combine modalities to broaden state-action coverage, tune temporal supervision for the execution rate, ablate state inputs judiciously, and invest in rig/lighting design.

Looking forward, we identify several priorities for future work. First, the intriguing finding that excluding proprioception can help calls for deeper diagnostic ablations to clarify underlying causes. Second, our evaluation was limited to two tasks; expanding task diversity will be essential for testing generalization to broader assistive settings. Third, while our dataset of 400 demonstrations enabled careful controlled comparisons, scaling to larger balanced collections, closer to benchmarks such as DROID or OXE, will better probe modality effects at scale. Fourth, the robustness of proprioception-related findings requires further exploration, ideally under matched protocols across joint-space and end-effector demonstrations. Fifth, environmental confounds such as rig setup and lighting conditions should be systematically isolated and mitigated to improve reproducibility. Finally, while Octo was an appropriate choice for controlled ablations, extending the analysis to larger-scale policies such as OpenVLA would test the generality of our conclusions.

Acknowledgments

This work is supported by the Aging in Place Challenge Program at the National Research Council of Canada (AiP-302).

References

[1] A. Brohan and et al. Rt-1: Robotics transformer for real-world control at scale. In *arXiv preprint arXiv:2212.06817*, 2022.

[2] A. Brohan and et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In J. Tan, M. Toussaint, and K. Darvish, editors, *Proceedings of the 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pages 2165–2183. PMLR, November 2023. URL https://proceedings.mlr.press/v229/zitkovich23a.html.

[3] D. Ghosh and et al. Octo: An open-source generalist robot policy. In *Proceedings of Robotics: Science and Systems (RSS)*, Delft, Netherlands, 2024. URL https://arxiv.org/abs/2405.12213. Accepted at RSS 2024.

[4] A. Khazatsky and et al. Droid: A large-scale in-the-wild robot manipulation dataset. In *Proceedings of Robotics: Science and Systems (RSS)*, 2024. URL https://arxiv.org/abs/2403.12945. arXiv preprint arXiv:2403.12945.

[5] A. O'Neill and et al. Open x-embodiment: Robotic learning datasets and rt-x models. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6892–6903, 2024. doi:10.1109/ICRA57147.2024.10611477. URL https://ieeexplore.ieee.org/document/10611477. Presented at ICRA 2024, Yokohama, Japan.

[6] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, pages 139–183. Elsevier, 1988.

[7] W. Zhang, Y. Zhou, D. Zhang, and et al. Effective tuning strategies for generalist robot manipulation policies. *arXiv preprint arXiv:2410.01220*, 2024. URL https://arxiv.org/abs/2410.01220.

[8] M. J. Kim and et al. Openvla: An open-source vision-language-action model. In P. Agrawal, O. Kroemer, and W. Burgard, editors, *Proceedings of the 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pages 2679–2713. PMLR, November 2025. doi:10.48550/arXiv.2406.09246. URL https://proceedings.mlr.press/v270/kim25c.html. Presented at CoRL 2024.

[9] K. Black and et al. π_0: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024. URL https://arxiv.org/abs/2410.24164.

[10] B. A. Newman, R. M. Aronson, S. S. Srinivasa, K. Kitani, and H. Admoni. Harmonic: A multimodal dataset of assistive human–robot collaboration. *The International Journal of Robotics Research*, 2021. doi:10.1177/02783649211050677. URL https://journals.sagepub.com/doi/10.1177/02783649211050677.

[11] R. Liu, A. Bhaskar, and P. Tokekar. Adaptive visual imitation learning for robotic assisted feeding across varied bowl configurations and food types, 2024. URL https://arxiv.org/abs/2403.12891.

[12] H. Li, Y. Cui, and D. Sadigh. How to train your robots? the impact of demonstration modality on imitation learning. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2025. URL https://arxiv.org/abs/2503.07017. Accepted for presentation.

[13] J. Yang, M. S. Mark, B. Vu, and et al. Robot fine-tuning made easy: Pre-training rewards and policies for autonomous real-world reinforcement learning. *arXiv preprint arXiv:2310.15145*, 2023. URL https://arxiv.org/abs/2310.15145.

[14] C. Chi, Z. Xu, S. Feng, and et al. Diffusion policy: Visuomotor policy learning via action diffusion. *arXiv preprint arXiv:2303.04137*, 2023. URL https://arxiv.org/abs/2303.04137.

[15] S. Ramos, S. Girgin, L. Hussenot, D. Vincent, H. Yakubovich, D. Toyama, A. Gergely, P. Stanczyk, R. Marinier, J. Harmsen, O. Pietquin, and N. Momchev. Rlds: an ecosystem to generate, share and use datasets in reinforcement learning, 2021.

[16] N. Gavenski, F. Meneguzzi, M. Luck, and O. Rodrigues. A survey of imitation learning methods, environments and metrics. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 37(4):Article 111, Aug. 2024. URL https://arxiv.org/abs/2404.19456.