Human-Gated Interactive Imitation Learning for Bimanual VR Teleoperation

A DAgger-style approach to closing the gap between offline accuracy and closed-loop success

Alex Wegener · Boston University · September 2026

Abstract. Behaviour cloning from teleoperated demonstrations gives manipulation policies that fit their training data well yet fail once they run closed-loop, because small errors carry the robot into states the demonstrations never visited. DAgger-style interactive imitation learning addresses this by collecting expert corrections in exactly those states. We describe a human-gated variant for a bimanual robot driven through an immersive VR interface: an action-chunking transformer policy runs autonomously, a remote operator takes over whenever it drifts toward failure, and the corrections are labelled, filtered and aggregated into the next training round. We summarise the objective, place the approach in the interactive imitation learning literature, and outline the open design questions and the planned evaluation.

1 Motivation

Our policy is an Action Chunking Transformer (ACT) [9] trained on roughly 60 teleoperated demonstrations of a bimanual sorting task. Offline it behaves as expected: validation error falls steadily and open-loop replay follows the demonstrated trajectories closely. Closed-loop it does not. The arms move purposefully toward the objects, then hesitate or stop a few centimetres from the grasp. The final approach is the rarest and most precise part of every demonstration, and it is exactly where the robot's own small errors produce observations the dataset does not contain.

This is the classic failure of behaviour cloning. The policy is trained on the state distribution induced by the expert but deployed on the distribution induced by itself, so errors compound over the horizon. Collecting more demonstrations helps only slowly, because new demonstrations again cover the expert's states rather than the policy's. What the policy needs is supervision *in the states it actually reaches*.

2 Background: DAgger and its human-gated variants

DAgger (Dataset Aggregation) [1] repeats a simple loop: run the current policy, have the expert label the states it visits, add those labels to the dataset, and retrain. Ross et al. show that this reduction to no-regret online learning bounds the error growth linearly in the task horizon, against quadratic growth for plain behaviour cloning. The catch for robotics is the labelling step: asking a human to provide the correct action for every state while the robot does something else is unnatural and imprecise.

Human-gated interactive imitation learning removes that catch. The robot acts on its own; the human watches and takes control only when it is about to fail, then hands control back. HG-DAgger [2] introduced this form and trains on the intervention segments. Later work asks who should decide when to intervene and how the resulting data should be weighted. Robot-gated methods let the policy request help from its own uncertainty or estimated risk, which reduces operator burden [3, 4, 5]. Weighting methods emphasise intervention samples over routine autonomous ones [6, 7]. Recent systems apply the idea to fine, contact-rich and bimanual manipulation with purpose-built teleoperation hardware [8].

Method	Who decides to intervene	What is trained on
DAgger [1]	Expert labels every visited state	All labels, aggregated over rounds
HG-DAgger [2]	Human, by taking over	Human intervention segments
EnsembleDAgger [3]	Robot, from ensemble disagreement	Expert labels on requested states
LazyDAgger / ThriftyDAgger [4, 5]	Robot, from novelty and risk, under a human budget	Interventions; fewer, longer handovers
IWR [6]	Human, by taking over	All data, interventions up-weighted
Sirius [7]	Human, by taking over	All data, reweighted by segment type (pre-intervention down-weighted)
RoboCopilot [8]	Human, seamless takeover in bimanual teleop	Interventions aggregated for fine manipulation
This work	Human, whole-robot takeover in VR	Human-authored chunks aggregated with demonstrations

Table 1. Interactive imitation learning methods relevant to this work. Numbers refer to the references.

3 Approach

The robot runs the current policy autonomously while an operator watches through the same stereo VR interface used for demonstration. Control is a small state machine with three values, POLICY, HUMAN and HOLD, owned by the robot side rather than by either requester, so the operator and the policy can never both command an arm.

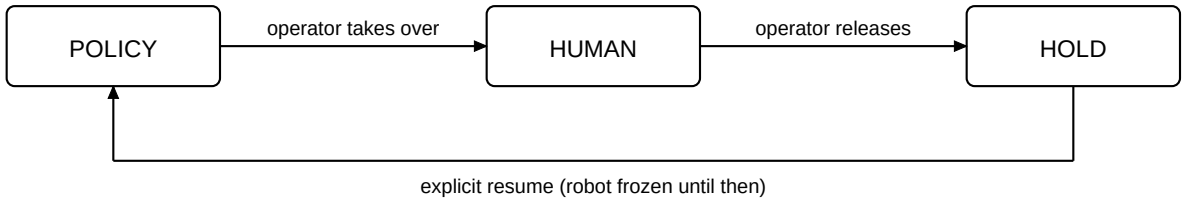


Figure 1. Authority states during a rollout. Release never hands control straight back to the policy.

Taking over. Pulling a controller trigger switches the whole robot, both arms and the head, to HUMAN. The operator's input is re-anchored to the pose the policy left behind, so the handover produces no jump, and any motion the policy had already planned is discarded immediately. This matters especially for chunked policies, which otherwise keep executing seconds of queued actions after the human has taken control.

Handing back. Releasing both triggers moves to HOLD, where the robot stops. The operator resumes the policy with an explicit gesture. Separating release from resume lets the operator re-clutch during a correction without the policy taking back an arm mid-repair.

Data. Every logged sample carries its authority label. The training set for the next round consists of the original demonstrations plus action chunks that were authored by the human for at least 90% of their horizon, so no training target mixes human and policy actions. The policy is retrained on the aggregate, and the loop repeats (Figure 2).

Operator feedback. The HUD shows authority, per-episode intervention count and autonomy share, inference time, and a coarse agreement indicator derived from how much the policy's overlapping chunk predictions disagree. The indicator is advisory only. It informs the operator's decision but never gates or blends control, keeping the system human-gated rather than shared control.

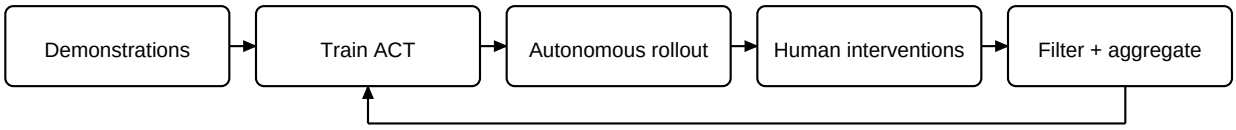


Figure 2. One DAgger round. Each round's rollouts are run with the policy trained in the previous round.

4 Open design questions

Whole-robot versus per-arm authority. Letting the operator correct one arm while the policy keeps the other is attractive for bimanual work and has precedent as traded control in classical telerobotics. The intervention learning literature, however, is almost entirely whole-embodiment. For a policy that predicts one joint action for both arms, a one-arm correction yields a label that is half the policy's own output. In our first sessions per-arm corrections produced no usable training data at all. We therefore use whole-robot authority for now. Per-arm correction with a loss masked to the corrected arm, i.e. partial supervision, remains a candidate contribution.

How to weight the data. HG-DAgger keeps only intervention segments; IWR and Sirius keep everything and reweight. With few demonstrations, discarding the autonomous segments wastes the successful behaviour they contain, while keeping them risks reinforcing the behaviour that led to intervention. We plan to compare the variants, which is a filter over one label array rather than a system change.

Operator burden and latency. The deployed setting is intercontinental teleoperation, so every takeover happens under network delay. Robot-gated triggers [4, 5] could reduce how often a person must watch closely; for now the disagreement signal is shown but not acted on.

5 Evaluation plan

Each round is scored by closed-loop task success on a fixed set of scene seeds, interventions per episode, and autonomy share (the fraction of engaged time under policy control), plotted against the cumulative amount of human data. The main comparison is against the same human effort spent on additional plain demonstrations: DAgger is worthwhile if interventions buy more success per operator-minute. Secondary measures are the handover itself (pose jump at takeover, which should be zero) and whether the data filter keeps a useful fraction of intervention time.

References

[1] S. Ross, G. Gordon, J. A. Bagnell. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. AISTATS, 2011.

[2] M. Kelly, C. Sidrane, K. Driggs-Campbell, M. J. Kochenderfer. HG-DAgger: Interactive Imitation Learning with Human Experts. ICRA, 2019.

[3] K. Menda, K. Driggs-Campbell, M. J. Kochenderfer. EnsembleDAgger: A Bayesian Approach to Safe Imitation Learning. IROS, 2019.

[4] R. Hoque, A. Balakrishna, C. Putterman, et al. LazyDAgger: Reducing Context Switching in Interactive Imitation Learning. CASE, 2021.

[5] R. Hoque, A. Balakrishna, E. Novoseller, et al. ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning. CoRL, 2021.

[6] A. Mandlekar, D. Xu, R. Martín-Martín, Y. Zhu, L. Fei-Fei, S. Savarese. Human-in-the-Loop Imitation Learning using Remote Teleoperation (IWR). arXiv:2012.06733, 2020.

[7] H. Liu, S. Nasiriany, L. Zhang, Z. Bao, Y. Zhu. Robot Learning on the Job: Human-in-the-Loop Autonomy and Learning During Deployment (Sirius). RSS, 2023; IJRR, 2025.

[8] P. Wu, Y. Shentu, Q. Liao, et al. RoboCopilot: Human-in-the-loop Interactive Imitation Learning for Robot Manipulation. arXiv:2503.07771, 2025.

[9] T. Z. Zhao, V. Kumar, S. Levine, C. Finn. Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware (ACT). RSS, 2023.

[10] C. Celemin, R. Pérez-Dattari, E. Chisari, et al. Interactive Imitation Learning in Robotics: A Survey. Foundations and Trends in Robotics, 2022.