MOTUS2: A SELF-EVOLVING GENERAL WORLD MODEL FOR DEXTEROUS MANIPULATION

Anonymous authors
Paper under double-blind review

ABSTRACT

General embodied agents should perceive, predict, act, evaluate, and improve within a unified system. World models have shown great promise in building such agents, yet existing models typically append an action output head to a world simulator, without coupling them into a closed decision-and-learning loop for policy improvement. We present **Motus2**, a self-evolving general world model for dexterous manipulation. Motus2 advances world modeling through model scaling and data scaling. For **model scaling**, a single model with shared weights exposes three control interfaces: a policy (*world–action model*), a simulator (*action-conditioned world model*), and an evaluator (*value model*). The policy proposes candidate action chunks, the simulator predicts their visual consequences, and the evaluator assesses the predicted outcomes. Their coupling forms a closed decision-and-learning loop for policy improvement. This formulation uses curated expert demonstrations for action learning, while failed and suboptimal interactions provide valuable evidence for dynamics modeling and value learning. For **data scaling**, Motus2 progresses from large-scale monocular egocentric data to synchronized stereo egocentric data, followed by robot-domain adaptation with robot trajectories and supplementary human–robot alignment data. Motus2 further studies global-autoregressive and hybrid-memory extensions of its sliding-window context, adds tactile feedback for contact-aware control, and is instantiated on a fully biomimetic platform with stereo vision, dual arms, dual dexterous hands, and tactile sensing. Together, egocentric data scaling and closed-loop general world model scaling provide a general path toward self-evolving dexterous manipulation.

1 INTRODUCTION

General-purpose dexterous manipulation requires a model to learn interaction priors, retain relevant history under self-occlusion, reason about contact, and improve beyond supervised imitation. Current robot foundation models rely largely on curated action-supervised demonstrations (Liu et al., 2025; 2026; Kim et al., 2025; Black et al., 2024; 2025; Feng et al., 2025; Chi et al., 2023). Such demonstrations are expensive to collect, and imitation alone does not evaluate an action's consequences or learn from unsuccessful outcomes.

Egocentric human data offers scalable interaction priors. Monocular recordings provide broad task and semantic coverage (Grauman et al., 2022; Yang et al., 2026b; Zheng et al., 2026b), while synchronized stereo observations add implicit depth cues and more accurate 3D hand-pose estimates. Transferring these priors to robots still requires embodiment adaptation, temporal context for occluded objects, and tactile feedback for contact events that are ambiguous from vision alone (Niu et al., 2026; Yang et al., 2026c; Hong et al., 2025).

Control-oriented world models provide two complementary interfaces: a *world–action model* proposes executable actions, and an *action-conditioned world model* predicts their consequences. Motus realizes both within one UniDiffuser-style video–action model (Bi et al., 2026a; Bao et al., 2023b). This follows the General World Model perspective that action grounds internal predictions in physical interaction (Zhu et al., 2026b). However, prediction alone does not determine whether an outcome advances a task. A third interface, an *evaluator*, must assess outcomes and supply feedback for policy improvement. The challenge is to realize these interfaces within shared param-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

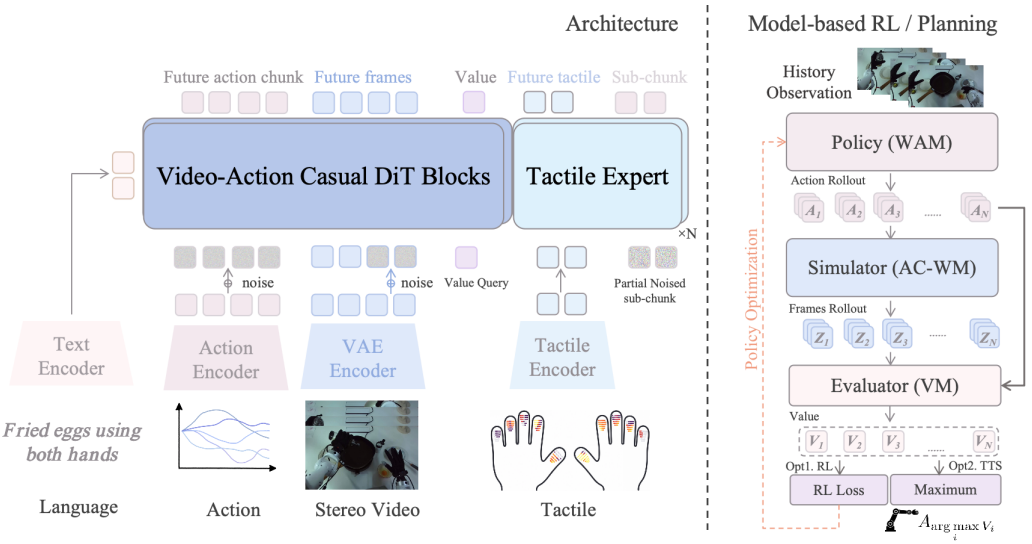


Figure 1: **Motus2 overview.** A single general world model with one shared parameter set exposes three control interfaces. The policy generates action chunks from language, current robot observations, and the selected visual working-memory context; the simulator predicts future visual states under candidate actions; and the evaluator assesses the resulting branches for planning and policy optimization.

eters without exposing future observations to action prediction, and to turn their scores into policy updates.

We present Motus2, a self-evolving General World Model for dexterous manipulation built on Motus. It scales along two axes. For *data scaling*, pretraining progresses from monocular to synchronized stereo egocentric data, followed by robot-domain adaptation with robot trajectories and supplementary human–robot alignment data. For *model scaling*, a shared model exposes a policy, simulator, and value evaluator through an action-first factorization. The policy proposes action chunks, the simulator predicts their visual consequences, and the evaluator estimates task progress. Best-of-N planning selects candidates at test time; model-based reinforcement learning (MBRL) with DiffusionNFT (Zheng et al., 2026a) converts their scores into policy updates. Curated successful demonstrations supervise action learning, while failed and suboptimal interactions supervise dynamics and outcome evaluation.

Our contributions are threefold: (1) a shared-parameter world model coupling action generation, future prediction, and task-progress evaluation, with long-history context and tactile extensions; (2) an egocentric data curriculum and robot-transfer study, including a stereo human-data scaling trend; and (3) value-guided planning and policy optimization that close the policy–simulator–evaluator loop. Detailed related work appears in Appendix A.

2 Method

2.1 Overview

Motus2 learns a shared video–action backbone through joint egocentric pre-training, then exposes policy, simulator, and evaluator interfaces through action-first factorization and trajectory-dependent supervision in robot-domain training (Fig. 1). Value-guided MBRL connects these interfaces into a policy-improvement loop; working memory and a tactile expert extend visual context and contact feedback.

We formulate language-conditioned dexterous manipulation as a POMDP (Appendix B.1). The observation $o_t = (I_t, q_t, \tau_t)$ contains vision, proprioception, and optional tactile sensing. Occlusion makes the current observation insufficient, so the policy uses history $h_t = (\ell, o_0, \ldots, o_t)$ under instruction ℓ. The core context c_t represents language, current proprioception, and encoded visual

history; the tactile expert separately uses the concurrent tactile stream. Let $A_t = (a_t, \ldots, a_{t+H-1})$ be an executable action chunk, Z_t its latent future observations, and Y_t a categorical task-progress value whose support comprises numerical bin centers. From robot-domain mid-training onward, we factor their density in the order used for control and planning:

$$p_\theta(A_t, Z_t, Y_t \mid c_t) = \underbrace{\pi_\theta(A_t \mid c_t)}_{\text{policy (WAM)}} \underbrace{p_\theta^{\text{wm}}(Z_t \mid c_t, A_t)}_{\text{simulator (AC-WM)}} \underbrace{p_\theta^{\text{vm}}(Y_t \mid c_t, A_t, Z_t)}_{\text{evaluator (VM)}}. \tag{1}$$

The factors are interfaces of the same shared-parameter model. Its branch score is the evaluator's conditional expectation:

$$V_\theta(c_t, A_t, Z_t) = \mathbb{E}_{Y_t \sim p_\theta^{\text{vm}}(\cdot \mid c_t, A_t, Z_t)}[Y_t]. \tag{2}$$

Ordinary control queries only $\pi_\theta(A_t \mid c_t)$; planning and policy optimization additionally materialize future video and value. Joint video–action training therefore does not require a future rollout at every control step (Ye et al., 2026a; Yuan et al., 2026).

2.2 General World Modeling

Stage-specific information flow. Pre-training learns joint video–action representations with mutual within-chunk visibility and no value query. At robot-domain mid-training, we introduce a read-only value query and an action-first mask. Each window contains clean teacher-forced observation history and chunk blocks comprising optional proprioception, actions, future video, and a value query. Action tokens cannot read their chunk's future-video or value tokens; future-video tokens may read the action; the value query reads both but is hidden from every other token. Across blocks, attention is causal and windowed: a chunk can access preceding clean observations but never a later block. This is *chunk-autoregressive* modeling: autoregression operates across chunks, while actions within each chunk are generated jointly by flow matching.

Both stages share a backbone initialized from a foundation video diffusion model (Team Wan et al., 2025). The mask changes the within-chunk dependency to $A_j \to Z_j \to U_j$, preserving tokenization, backbone, and cross-chunk causality. Stereo latents share temporal and vertical coordinates but use distinct horizontal RoPE ranges. Proprioception, actions, and value queries enter the same transformer; language enters through cross-attention. Figure 2 compares the masks; Appendix B.1 gives the formal token layout.

Trajectory-dependent supervision routing. A causal mask cannot determine whether a recorded action is desirable. Failed and suboptimal executions provide transition and outcome evidence, but should not become imitation targets. For a clean target $x \in \{Z_j, A_j\}$, Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$, and noise level $\sigma \in [0, 1]$, we use

$$x_\sigma = (1 - \sigma)x + \sigma\epsilon, \qquad v^* = \epsilon - x. \tag{3}$$

The shared core predicts video and action velocities v_θ^z, v_θ^a and the categorical value distribution p_θ^{vm}, trained with

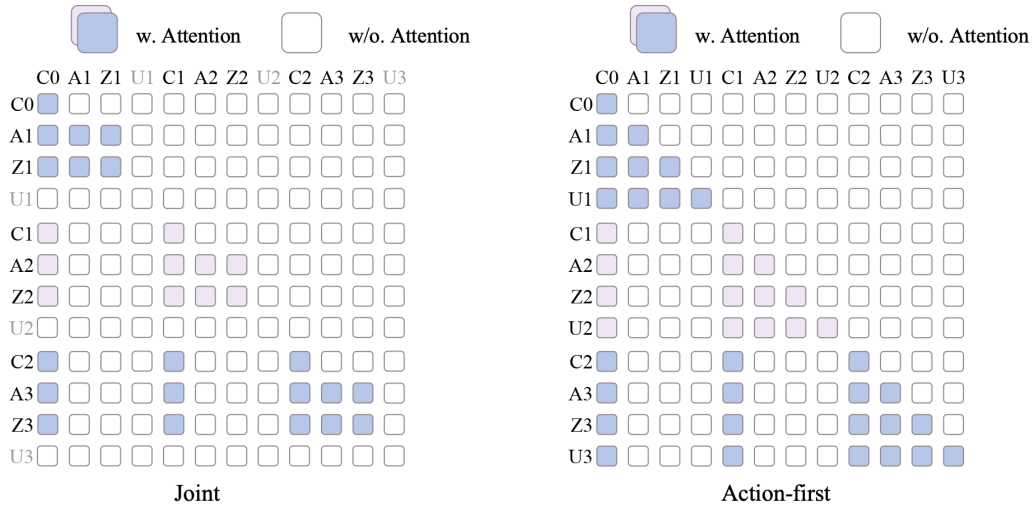
$$\mathcal{L} = w_z \|v_\theta^z - v_z^*\|_2^2 + \lambda_a w_a \|v_\theta^a - v_a^*\|_2^2 + \lambda_v w_v \, \text{CE}(p_\theta^{\text{vm}}, Y_t). \tag{4}$$

Independent video/action noise levels and loss gates define three modes:

mode	(σ^z, σ^a)	(w_z, w_a, w_v)	supervised conditional
policy	$(> 0, > 0)$	$(1, 1, 0)$	$p_\theta(A_t, Z_t \mid c_t)$
simulation	$(> 0, 0)$	$(1, 0, 0)$	$p_\theta(Z_t \mid c_t, A_t)$
evaluation	$(0, 0)$	$(0, 0, 1)$	$p_\theta(Y_t \mid c_t, A_t, Z_t)$

(5)

Policy mode jointly supervises action and video under the first two factors of Eq. (1); deployment reads only its action factor. Simulation conditions on clean recorded actions, while evaluation conditions on both clean actions and video. From mid-training onward, only curated successful trajectories activate action supervision. Failed, suboptimal, and task-irrelevant interactions are routed to applicable simulation or evaluation modes, where actions remain conditioning variables. Thus their transitions and outcomes improve world modeling without teaching the policy to reproduce undesirable behavior.



(a) Joint mask for pre-training.

(b) Action-first mask for mid or post-training.

Figure 2: **Stage-specific chunk masks.** Context (C), video (Z), and action (A) tokens share one backbone; a read-only value query (U) is introduced from mid-training onward and is inactive during joint pre-training. Joint pre-training permits bidirectional video–action interaction within a chunk. The action-first mask blocks current future-video information from action prediction, while future-video tokens may read the action and the value query may read both. Both masks remain causal across chunks.

Data-scaling curriculum. Pre-training Stage 1 learns the video pathway from low- then high-resolution monocular egocentric clips with bidirectional conditional flow matching and one or two clean latent context frames. Stage 2 jointly trains synchronized stereo views and action tokens with bidirectional within-chunk visibility. Mid-training grounds these human-interaction priors in robot trajectories, supplemented by human–robot alignment data, and introduces the action-first mask and three-mode supervision routing. Post-training retains these interfaces for target-robot SFT, value-guided MBRL, long-history context, and tactile feedback. The ordering matters: action-first dependencies and trajectory-dependent action supervision are introduced during robot-domain adaptation, after joint human-interaction representation learning.

2.3 Value-Guided Closed-Loop Self-Evolution

The evaluator supplies feedback for both selecting actions and improving the policy. Because all three interfaces share one model, Motus2 evaluates its own predicted consequences to improve its action pathway. The complete loop is illustrated in Appendix B.2.

2.3.1 Progress-Based Value Learning

Following VLAC's relative-progress formulation (Zhang et al., 2026a), successful trajectory segments receive

$$r_t = \frac{\Delta t}{T - t}, \tag{6}$$

where Δt is segment length, T the trajectory length, and t its starting timestep. This target lies in $[0, 1]$ and measures segment progress relative to the remaining horizon. To teach the evaluator how to score deviations, failed teleoperation trajectories supply realistic manipulation errors and unintended contacts, while task-irrelevant interactions supply behavior unrelated to the instruction. Each negative segment receives

$$r_t = -\frac{\Delta t}{T - t}. \tag{7}$$

Positive and negative rewards are discretized into Y_t for the categorical value loss in Eq. (4). These targets give temporally localized feedback at different stages of execution.

2.3.2 Planning and Policy Optimization

Model-based planning. At inference, we propose N action chunks, simulate their consequences Z_i, and score each branch with $V_i = V_\theta(c_t, A_i, Z_i)$. Best-of-$N$ selects $i^\star = \arg\max_i V_i$ and executes $A_{i^\star}$. We then append the real observation and replan, limiting reliance on long imagined rollouts.

Policy optimization. Planning selects among current proposals; MBRL changes their distribution. We use DiffusionNFT (Zheng et al., 2026a) to convert branch scores into flow-matching action updates. Each generated action is noised as in Eq. (3), giving target $v_i^{\text{tar}} = \epsilon_i - A_i$. The online policy and EMA reference predict $v_{\theta,i}$ and $v_{\text{ref},i}$ from the same noised action, noise level, instruction, robot state, and clean observation history. The simulated future Z_i affects the update only through V_i; the action predictor never reads it.

We standardize generated-candidate values using their population standard deviation with an ε floor, clip to $[-1, 1]$, and map to optimality weights $\hat{r}_i \in [0, 1]$. Each group additionally contains one ground-truth action–future pair with $\hat{r}_i = 1$, excluded from the normalization statistics. With $\beta > 0$, define implicit fields $v_i^+ = (1-\beta)v_{\text{ref},i} + \beta v_{\theta,i}$ and $v_i^- = (1+\beta)v_{\text{ref},i} - \beta v_{\theta,i}$, keeping the reference prediction fixed during each update. The core objective is

$$\mathcal{L}_{\text{NFT}} = \mathbb{E}_i\left[\hat{r}_i \|v_i^+ - v_i^{\text{tar}}\|_2^2 + (1-\hat{r}_i)\|v_i^- - v_i^{\text{tar}}\|_2^2\right]. \tag{8}$$

High-value candidates emphasize the positive field and low-value candidates the negative field; weights near 0.5 keep the policy near its reference. Only action-related parameters are updated; the shared video backbone and evaluator remain frozen. A Ray-based asynchronous pipeline decouples rollout, scoring, reference-target construction, and optimization, using versioned EMA references and bounded sample staleness. Appendix B.2 gives the full formulation, and Appendix B.3 details the pipeline.

2.4 Working Memory

Occlusion and delayed use of earlier evidence motivate explicit history retention in c_t. Our default sliding window caches recent clean observed latents, appends observations after execution, and evicts the oldest when full; window-relative temporal RoPE is rebased after eviction. This bounds cache capacity and per-step attention cost. Two post-trained extensions retain longer histories: global autoregression keeps all preceding observations with episode-level positions, while MemoryWAM-style hybrid memory (Yang et al., 2026c) keeps initial anchors and a recent window at full resolution and older intermediate observations as persistent memory tokens. The former's cost grows with episode length; the latter compresses older evidence. Both train on complete episodes packed by latent-frame load with isolated causal attention. Appendix B.4 gives cache and packing details; §4.5 compares the extensions.

2.5 Tactile Expert: Tactile Refinement and Prediction

Vision can leave contact ambiguous even with longer history. Following T-Rex (Niu et al., 2026), a lightweight tactile expert reuses the backbone's intermediate action chunk and detached layer-wise KV cache. The backbone once denoises the full chunk from $\sigma = 1$ to σ_c; immediately before each short sub-chunk executes, the expert uses the latest tactile window to complete the $\sigma_c \to 0$ refinement, without rerunning the backbone. An auxiliary flow-matching loss predicts the following real force window in normalized signal space. Future-force tokens may read intermediate actions, but action-refinement tokens cannot read noised future forces. Sharpa Wave deformation images serve only as clean conditioning inputs. The expert is trained with action-refinement and force-prediction losses; deployment outputs only refined actions. Appendix B.5 provides the full objectives and rolling-update details.

3 Large-Scale Egocentric Dataset

We construct an approximately 130K-hour egocentric corpus from open-source and procured monocular and stereo recordings. Monocular data provides broad visual and semantic coverage;

synchronized stereo data adds implicit depth cues and more accurate 3D hand-pose estimates. These are raw recording hours before quality filtering, segmentation, and annotation. Source-level amounts and acquisition modes are reported in Appendix C.

Pretraining uses low-resolution monocular clips for 500K steps, high-resolution monocular clips for 340K steps, and stereo egocentric data for 450K steps. Robot-domain mid-training then uses more than 100 hours of robot and human–robot alignment data; the latter includes tens of hours collected with Wuji Human Gloves. Processing standardizes hand poses to a 134-D representation and applies synchronized quality-aware temporal segmentation and vision–language annotation. The corpus spans diverse manipulation primitives, objects, tasks, and scenes. Collection, processing, and training details are in Appendices E, F, and H.

4 EXPERIMENTS

In this section, we aim to answer the following research questions through our experiments:
RQ1: *How does egocentric human pretraining affect downstream robot control?*
RQ2: *Do MBRL and test-time planning improve performance beyond supervised imitation?*
RQ3: *How do different context mechanisms affect long-horizon control?*
RQ4: *Does the tactile expert improve contact-rich manipulation?*

4.1 EXPERIMENT SETUP

Task suites. The main suite contains five target-robot tasks spanning diverse control requirements: Place Ball and Put Phone test spatial grounding and placement; Attach Eraser and Screw Bulb require precise object alignment and sustained contact; and Multi-Finger tests coordinated multi-finger grasping and manipulation of differently shaped objects. The MBRL study uses Put Phone and Multi-Finger to isolate the effects of policy optimization and test-time planning on the same physical tasks. The memory study uses the separate Find Square and Press Button probes, whose relevant evidence may be occluded or delayed beyond a short observation window. This separation prevents the broad manipulation suite from obscuring the specific effect of persistent context. The tactile study uses Pull Out the Paper Cup and Tear Paper, which require contact-sensitive grasp adjustment and sustained bimanual interaction.

Evaluation protocol. Within each evaluation setting, all methods use the same initial configurations for each task. Real-robot evaluations use 20 configurations per task, whereas the simulation branch of the long-horizon context study uses 25. Each method executes one closed-loop rollout per configuration, using identical language instructions, episode horizons, and task-specific success checkers. A rollout is counted as successful only when all required task predicates are satisfied within the horizon; partial completion is counted as failure. We report success rate (SR), with the number of rollouts specified for each study. The average is a macro-average over tasks, so each task receives equal weight. Simulation and real-robot results are reported separately.

Controlled comparisons. We compare Motus2 against $\pi_{0.5}$ (Black et al., 2025) under matched target-task SFT data, observation interface, and evaluation protocol. WAN-SFT, Pretrain-SFT, and Motus2 (Midtrain-SFT) use the same SFT procedure and differ only in initialization: WAN, the two-stage egocentric-pretraining checkpoint, and the robot-domain mid-trained checkpoint, respectively. Within each controlled study, all other factors are held fixed.

4.2 MAIN RESULTS

Results. Egocentric pretraining raises the macro-average success rate from 0% for WAN-SFT to 51% for Pretrain-SFT. Motus2 (Midtrain-SFT) further increases the average to 84%, a 33-point gain over Pretrain-SFT. Motus2 reaches 100% on Place Ball and Attach Eraser, 90% on Screw Bulb, 70% on Multi-Finger, and 60% on Put Phone. Under the matched target-task SFT protocol, these comparisons support the contributions of egocentric pretraining and subsequent robot-domain mid-training.

Table 1: **Main results.** Each entry reports success rate over 20 rollouts under matched target-robot finetuning. WAN-SFT, Pretrain-SFT, and Midtrain-SFT differ only in initialization.

Method	Place Ball	Multi-Finger	Attach Eraser	Screw Bulb	Put Phone	Avg. SR ↑
$\pi_{0.5}$	0%	0%	0%	0%	0%	0%
WAN-SFT	0%	0%	0%	0%	0%	0%
Pretrain-SFT	60%	35%	90%	55%	15%	51%
Motus2 (Midtrain-SFT)	**100%**	**70%**	**100%**	**90%**	**60%**	**84%**

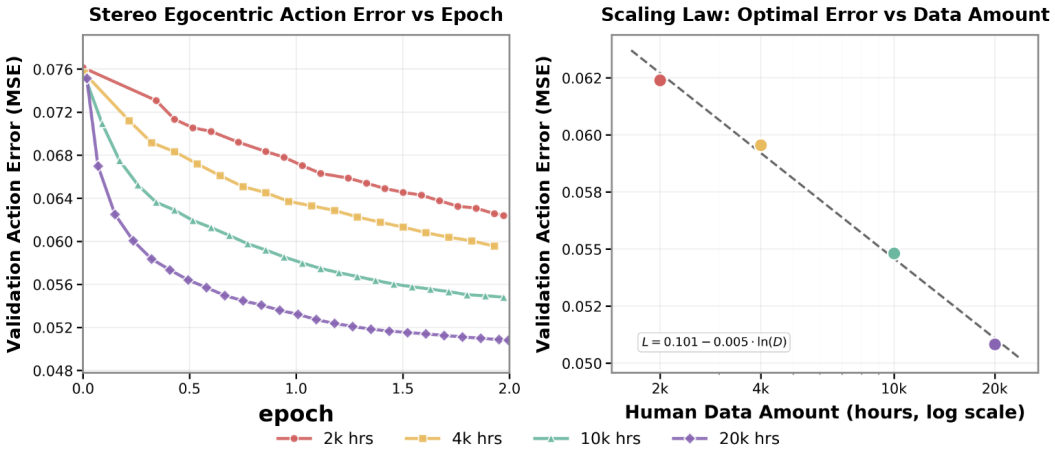


Figure 3: **Scaling laws of stereo egocentric human data.** Left: held-out human-action prediction error across training epochs for models trained on subsets constructed from 2k, 4k, 10k, and 20k raw recording hours of stereo data. Right: optimal validation error versus raw stereo human-data hours on a logarithmic scale, together with the fitted log-linear scaling law. The reported hours are measured before quality filtering and temporal segmentation; the effective amount of processed training data is therefore smaller.

4.3 Scaling Laws of Stereo Egocentric Human Data

Prior work has reported action-prediction scaling trends with monocular egocentric human data (Zheng et al., 2026b; Dyna Robotics, 2026). We study the corresponding trend for synchronized stereo observations and dexterous human actions using nested 2k, 4k, 10k, and 20k-hour subsets. The subsets preserve the source and resolution proportions, and all checkpoints are evaluated on the same trajectory-disjoint stereo held-out set.

Held-out human-action prediction error serves as the measure of generalization. For each validation sample, the flow-matching policy generates multiple action chunks. Their mean prediction is compared with the normalized ground-truth wrist and hand actions using mean squared error. Let $\mathcal{L}_{\text{val}}(D, e)$ denote the validation error after epoch e for a model pretrained on D hours of data. The optimal validation error at each data scale is

$$\mathcal{L}^*_{\text{val}}(D) := \min_e \mathcal{L}_{\text{val}}(D, e). \tag{9}$$

Fitting the optimal error against data hours yields the log-linear relation

$$\mathcal{L}^*_{\text{val}}(D) \approx 0.101 - 0.005 \ln D. \tag{10}$$

Figure 3 presents the training dynamics and optimal validation errors. As the pretraining corpus grows from 2k to 20k hours, the validation curves shift consistently downward: smaller datasets plateau earlier, whereas larger datasets attain lower errors at comparable epochs. Across the four data scales, the optimal validation error decreases monotonically and is approximately linear in

the logarithm of corpus size, indicating that the scaling trend previously observed for monocular egocentric data also extends to stereo observations within the measured range.

4.4 Model-Based Policy Improvement and Test-Time Planning

Factorial design. Starting from the same target-robot checkpoint, we independently enable DiffusionNFT policy optimization and Best-of-N planning. This yields a 2×2 comparison of Motus2, Motus2 + Planning, Motus2 + MBRL, and Motus2 + MBRL + Planning.

Table 2: **MBRL and planning results.** Each entry reports success rate over 20 rollouts. All variants start from the same target-robot checkpoint. MBRL updates the policy weights, whereas planning changes only inference.

Method	Put Phone	Multi-Finger	Avg. SR ↑
Motus2	60%	70%	65.0%
Motus2 + Planning	65%	70%	67.5%
Motus2 + MBRL	65%	80%	72.5%
Motus2 + MBRL + Planning	**70%**	**80%**	**75.0%**

Results. Planning raises Motus2 from 65.0% to 67.5% average success, showing a 2.5-point gain from online candidate selection without changing the policy weights. Motus2 + MBRL reaches 72.5% under direct inference, a 7.5-point gain over Motus2. Combining MBRL and planning yields 75.0%. Planning therefore remains beneficial after MBRL (2.5 points), while MBRL remains beneficial with planning enabled (7.5 points), showing observed gains from policy optimization and test-time selection in the evaluated tasks.

Qualitative value trajectories. Appendix I shows paired successful and failed executions for Cut Reeds, Make Tea, Multi-Finger, and Put Phone. The examples illustrate how predicted progress changes during execution and after failures; these scores are not calibrated success probabilities.

4.5 Long-Horizon Context Mechanisms

Protocol. On the Find Square and Press Button probes, we compare global autoregression and hybrid working memory as described in §2.4. The two variants use matched training and paired evaluation configurations in simulation and on the real robot.

Table 3: **Evaluation of long-horizon context mechanisms.** Simulation success rates are computed over 25 rollouts and real-robot rates over 20. Tasks are the long-horizon memory probes Find Square (`find_the_square`) and Press Button (`press_button`).

Setting	Context mechanism	Find Square	Press Button	Avg. SR ↑
Simulation	Hybrid memory	64%	40%	52%
	Global autoregression	84%	72%	78%
Real robot	Hybrid memory	30%	20%	25.0%
	Global autoregression	65%	50%	57.5%

Results. Global autoregression outperforms hybrid memory on both probes in simulation, reaching 84% on Find Square and 72% on Press Button, compared with 64% and 40% for hybrid memory. Its simulation macro-average is therefore 78%, versus 52% for hybrid memory. The same ordering holds on the real robot: global autoregression achieves 65% and 50% on the two tasks, whereas hybrid memory reaches 30% and 20%, giving macro-averages of 57.5% and 25%, respectively. Global autoregression performs better than hybrid memory in both simulation and real-robot evaluations.

4.6 Effect of Tactile Feedback

Protocol. Both variants start from the same mid-trained checkpoint and use the same target-task post-training data. The w/o Tactile variant removes the tactile expert, whereas the w/ Tactile variant uses the complete refinement and prediction objective in §2.5. We evaluate both variants on Pull Out the Paper Cup with the Sharpa Wave and Tear Paper with Wuji Hand 2.

Table 4: **Ablation of tactile feedback from the same mid-trained checkpoint.** Each entry reports the success rate over 20 rollouts.

Variant	Pull Out Paper Cup	Tear Paper	Avg. SR ↑
w/o Tactile	65%	55%	60.0%
w/ Tactile	75%	70%	72.5%

Results. The tactile expert improves success on Pull Out the Paper Cup from 65% to 75% and on Tear Paper from 55% to 70%. The macro-average increases from 60.0% to 72.5%, a 12.5-point gain. These results indicate that tactile feedback is beneficial for fine-grained, contact-sensitive manipulation in the evaluated tasks.

5 Conclusion

Motus2 couples action generation, action-conditioned simulation, and outcome evaluation within a shared General World Model. A curriculum from egocentric human data to robot-domain adaptation supplies interaction priors, while trajectory-dependent supervision uses successful demonstrations for imitation and unsuccessful interactions for dynamics and value learning. The resulting interfaces support both test-time planning and model-based policy improvement; long-history context and tactile feedback address complementary challenges in dexterous execution.

Limitations. Wearable tactile signals remain noisy under glove deformation, and differences between human and robot hand morphology limit direct cross-embodiment tactile transfer. These constraints motivate better wearable sensors and more anthropomorphic hands; further discussion is in Appendix K.

AI USE STATEMENT

Generative AI was used to assist with manuscript restructuring, language editing, LaTeX formatting, and anonymization (including face masking in an equipment photograph) for this submission. The authors are responsible for the final content of the manuscript.

REFERENCES

Fan Bao, Chongxuan Li, Jun Zhu, and Bo Zhang. Analytic-DPM: An analytic estimate of the optimal reverse variance in diffusion probabilistic models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=0xiJLKH-ufZ.

Fan Bao, Shen Nie, Kaiwen Xue, Yue Cao, Chongxuan Li, Hang Su, and Jun Zhu. All are worth words: A vit backbone for diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22669–22679, 2023a.

Fan Bao, Shen Nie, Kaiwen Xue, Chongxuan Li, Shi Pu, Yaole Wang, Gang Yue, Yue Cao, Hang Su, and Jun Zhu. One transformer fits all distributions in multi-modal diffusion at scale. In *International Conference on Machine Learning*, pp. 1692–1717, 2023b.

Fan Bao, Chendong Xiang, Gang Yue, Guande He, Hongzhou Zhu, Kaiwen Zheng, Min Zhao, Shilong Liu, Yaole Wang, and Jun Zhu. Vidu: a highly consistent, dynamic and skilled text-to-video generator with diffusion models. *arXiv preprint arXiv:2405.04233*, 2024.

Hongzhe Bi, Hengkai Tan, Shenghao Xie, Zeyuan Wang, Shuhe Huang, Haitian Liu, Ruowen Zhao, Yao Feng, Chendong Xiang, Yinze Rong, Hongyan Zhao, Hanyu Liu, Zhizhong Su, Lei Ma, Hang Su, and Jun Zhu. Motus: A unified latent action world model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 35101–35113, June 2026a. URL https://openaccess.thecvf.com/content/CVPR2026/html/Bi_Motus_A_Unified_Latent_Action_World_Model_CVPR_2026_paper.html.

Hongzhe Bi, Lingxuan Wu, Tianwei Lin, Hengkai Tan, Zhizhong Su, Hang Su, and Jun Zhu. H-rdt: Human manipulation enhanced bimanual robotic manipulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026b.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lindsay Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Lucy Xiaoyang Shi, James Tanner, Quan Vuong, Anna Walling, Huamin Wang, and Ury Zhilinsky. 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.

Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Robert Equi, Chelsea Finn, Niccolo Fusai, Manuel Y Galliker, et al. $\pi_{0.5}$: A vision-language-action model with open-world generalization. In *Proceedings of the 9th Conference on Robot Learning*, volume 305, pp. 17–40, 2025. URL https://proceedings.mlr.press/v305/black25a.html.

Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. Video generation models as world simulators. 2024.

Jake Bruce, Michael D Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, et al. Genie: Generative interactive environments. In *Forty-first International Conference on Machine Learning*, 2024.

Huayu Chen, Cheng Lu, Chengyang Ying, Hang Su, and Jun Zhu. Offline reinforcement learning via high-fidelity generative behavior modeling. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.14548.

Huayu Chen, Cheng Lu, Zhengyi Wang, Hang Su, and Jun Zhu. Score regularized policy optimization through diffusion behavior. In *International Conference on Learning Representations*, 2024a. URL https://arxiv.org/abs/2310.07297.

Huayu Chen, Kaiwen Zheng, Hang Su, and Jun Zhu. Aligning diffusion behaviors with q-functions for efficient continuous control. In *Advances in Neural Information Processing Systems*, volume 37, 2024b. URL https://arxiv.org/abs/2407.09024.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Robotics: Science and Systems*, 2023. URL https://diffusion-policy.cs.columbia.edu/.

Dyna Robotics. Dyna-2: A 1-million-hour scaling law for world-action models. Technical report, August 2026. URL https://dyna.co/dyna-2.

Yao Feng, Hengkai Tan, Xinyi Mao, Chendong Xiang, Guodong Liu, Shuhe Huang, Hang Su, and Jun Zhu. ViDAR: Embodied video diffusion model for generalist manipulation, 2025. URL https://arxiv.org/abs/2507.12898.

Yankai Fu, Ning Chen, Junkai Zhao, Shaozhe Shan, Guocai Yao, Pengwei Wang, Zhongyuan Wang, and Shanghang Zhang. Metis: Multi-source egocentric training for integrated dexterous vision-language-action model. *arXiv preprint arXiv:2511.17366*, 2025. URL https://arxiv.org/abs/2511.17366.

Shenyuan Gao, William Liang, Kaiyuan Zheng, Ayaan Malik, Seonghyeon Ye, Sihyun Yu, Wei-Cheng Tseng, Yuzhu Dong, Kaichun Mo, Chen-Hsuan Lin, Qianli Ma, Seungjun Nah, Loic Magne, Jiannan Xiang, Yuqi Xie, Ruijie Zheng, Dantong Niu, You Liang Tan, K. R. Zentner, George Kurian, Suneel Indupuru, Pooya Jannaty, Jinwei Gu, Jun Zhang, Jitendra Malik, Pieter Abbeel, Ming-Yu Liu, Yuke Zhu, Joel Jang, and Linxi Fan. DreamDojo: A generalist robot world model from large-scale human videos. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306, 2026. URL https://rpl.cs.utexas.edu/publications/2026/07/19/gao-icml26-dreamdojo/. Spotlight.

GigaWorld Team, Angyuan Ma, Boyuan Wang, Bohan Li, Chaojun Ni, Guo Li, Guan Huang, Guosheng Zhao, Hao Li, Hengtao Li, Jingyu Liu, Jiwen Lu, Qiuping Deng, Tingdong Yu, Xuancheng Xu, Xinyu Zhou, Xiuwei Xu, Xinze Chen, Xiaofeng Wang, Xiaoyu Tian, Yang Wang, Yifan Chang, Yukun Zhou, Yun Ye, Zhenyu Wu, Zhanqian Wu, and Zheng Zhu. GigaWorld-1: A roadmap to build world models for robot policy evaluation. *arXiv preprint arXiv:2607.02642*, 2026. URL https://arxiv.org/abs/2607.02642.

Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, Miguel Martin, Tushar Nagarajan, Ilija Radosavovic, Santhosh K. Ramakrishnan, Fiona Ryan, Jayant Sharma, Michael Wray, Mengmeng Xu, Eric Zhongcong Xu, Chen Zhao, Siddhant Bansal, et al. Ego4D: Around the world in 3,000 hours of egocentric video. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18995–19012, 2022.

Yanjiang Guo, Lucy Xiaoyang Shi, Jianyu Chen, and Chelsea Finn. Ctrl-World: A controllable generative world model for robot manipulation. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/0ae94013da7cd459402fd77874e09ee3-Paper-Conference.pdf.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023. URL https://arxiv.org/abs/2301.04104.

Yicong Hong, Yiqun Mei, Chongjian Ge, Yiran Xu, Yang Zhou, Sai Bi, Yannick Hold-Geoffroy, Mike Roberts, Matthew Fisher, Eli Shechtman, et al. Relic: Interactive video world model with long-horizon memory. *arXiv preprint arXiv:2512.04040*, 2025.

Chia-Yu Hung, Navonil Majumder, Haoyuan Deng, Liu Renhang, Yankang Ang, Amir Zadeh, Chuan Li, Dorien Herremans, Ziwei Wang, and Soujanya Poria. Nora-1.5: A vision-language-action model trained using world model- and action-based preference rewards. *arXiv preprint arXiv:2511.14659*, 2025. URL https://arxiv.org/abs/2511.14659.

Zhenyu Jiang, Yuqi Xie, Kevin Lin, Zhenjia Xu, Weikang Wan, Ajay Mandlekar, Linxi Fan, and Yuke Zhu. Dexmimicgen: Automated data generation for bimanual dexterous manipulation via imitation learning. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025. URL https://arxiv.org/abs/2410.24185.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. OpenVLA: An open-source vision-language-action model. In *Proceedings of the 8th Conference on Robot Learning*, volume 270, pp. 2679–2713, 2025. URL https://proceedings.mlr.press/v270/kim25c.html.

Baoyu Li, Xinchen Yin, Mengying Lin, Yixin Zhang, and Danfei Xu. Egowam: World action models beyond pixels with in-the-wild egocentric human data. *arXiv preprint arXiv:2607.08436*, 2026. URL https://arxiv.org/abs/2607.08436.

Songming Liu, Lingxuan Wu, Bangguo Li, Hengkai Tan, Huayu Chen, Zhengyi Wang, Ke Xu, Hang Su, and Jun Zhu. RDT-1B: a diffusion foundation model for bimanual manipulation. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/49f80e4d2471ad4f2edf4f5f1ab62339-Paper-Conference.pdf.

Songming Liu, Bangguo Li, Kai Ma, Lingxuan Wu, Hengkai Tan, Xiao Ouyang, Hang Su, and Jun Zhu. RDT2: Exploring the scaling limit of umi data towards zero-shot cross-embodiment generalization. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=ABrWtqsJYp.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in neural information processing systems*, 35:5775–5787, 2022.

Cheng Lu, Huayu Chen, Jianfei Chen, Hang Su, Chongxuan Li, and Jun Zhu. Contrastive energy prediction for exact energy-guided diffusion sampling in offline reinforcement learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22825–22855, 2023. URL https://proceedings.mlr.press/v202/lu23d.html.

Chengxuan Lu, Shukuan Wang, Yanjie Li, Yingying Fang, Huoyan Wang, Tian Zhang, Wei Liu, Shiji Jin, Fuyuan Qian, Peiming Li, Chao Xu, Baigui Sun, and Yang Liu. Accerl: A distributed asynchronous reinforcement learning and world model framework for vision-language-action models. *arXiv preprint arXiv:2603.18464*, 2026. URL https://arxiv.org/abs/2603.18464.

Hao Luo, Wanpeng Zhang, Yicheng Feng, Sipeng Zheng, Haiweng Xu, Chaoyi Xu, Ziheng Xi, Yuhui Fu, and Zongqing Lu. Being-H0.7: A latent world-action model from egocentric videos. *arXiv preprint arXiv:2605.00078*, 2026. URL https://arxiv.org/abs/2605.00078.

Motubrain Team, Chendong Xiang, Fan Bao, Haitian Liu, Hengkai Tan, Hongzhe Bi, James Li, Jiabao Liu, Jingrui Pang, Kiro Jing, Louis Liu, Mengchen Cai, Rongxu Cui, Ruowen Zhao, Runqing Wang, Shuhe Huang, Yao Feng, Yinze Rong, Zeyuan Wang, and Jun Zhu. MotuBrain: An advanced world action model for robot control. *arXiv preprint arXiv:2604.27792*, 2026. URL https://arxiv.org/abs/2604.27792.

Dantong Niu, Zhuoyang Liu, Zekai Wang, Boning Shao, Zhao-Heng Yin, Anirudh Pai, Yuvan Sharma, Stefano Saravalle, Ruijie Zheng, Jing Wang, et al. T-rex: Tactile-reactive dexterous manipulation. *arXiv preprint arXiv:2606.17055*, 2026.

Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, et al. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

Zhihao Wang, Jianxiong Li, Yu Cui, Yuan Gao, Xianyuan Zhan, Junzhi Yu, and Xiao Ma. World value models for robotic manipulation, 2026.

Jiazhi Yang, Kunyang Lin, Jinwei Li, Wencong Zhang, Tianwei Lin, Longyan Wu, Zhizhong Su, Hao Zhao, Ya-Qin Zhang, Li Chen, Ping Luo, Xiangyu Yue, and Hongyang Li. RISE: Self-improving robot policy with compositional world model. In *Robotics: Science and Systems*, 2026a. URL https://roboticsconference.org/program/papers/12/.

Ruihan Yang, Qinxi Yu, Yecheng Wu, Rui Yan, Borui Li, An-Chieh Cheng, Xueyan Zou, Yunhao Fang, Xuxin Cheng, Ri-Zhao Qiu, Hongxu Yin, Sifei Liu, Song Han, Yao Lu, and Xiaolong Wang. EgoVLA: Learning vision-language-action models from egocentric human videos. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2026b. URL https://rchalyang.github.io/EgoVLA/.

Sizhe Yang, Juncheng Mu, Tianming Wei, Chenhao Lu, Xiaofan Li, Linning Xu, Zhengrong Xue, Zhecheng Yuan, Dahua Lin, Jiangmiao Pang, and Huazhe Xu. MemoryWAM: Efficient world action modeling with persistent memory. *arXiv preprint arXiv:2606.20562*, 2026c.

Angen Ye, Boyuan Wang, Chaojun Ni, Guan Huang, Guosheng Zhao, Hao Li, Hengtao Li, Jie Li, Jindi Lv, Jingyu Liu, Min Cao, Peng Li, Qiuping Deng, Wenjun Mei, Xiaofeng Wang, Xinze Chen, Xinyu Zhou, Yang Wang, Yifan Chang, Yifan Li, Yukun Zhou, Yun Ye, Zhichao Liu, and Zheng Zhu. GigaWorld-Policy: An efficient action-centered world–action model. *arXiv preprint arXiv:2603.17240*, 2026a. URL https://arxiv.org/abs/2603.17240.

Seonghyeon Ye, Yunhao Ge, Kaiyuan Zheng, Shenyuan Gao, Sihyun Yu, George Kurian, Suneel Indupuru, You Liang Tan, Chuning Zhu, Jiannan Xiang, Ayaan Malik, Kyungmin Lee, William Liang, Nadun Ranawaka, Jiasheng Gu, Yinzhen Xu, Guanzhi Wang, Fengyuan Hu, Avnish Narayan, Johan Bjorck, Jing Wang, Gwanghyun Kim, Dantong Niu, Ruijie Zheng, Yuqi Xie, Jimmy Wu, Qi Wang, Ryan Julian, Danfei Xu, Yilun Du, Yevgen Chebotar, Scott Reed, Jan Kautz, Yuke Zhu, Linxi "Jim" Fan, and Joel Jang. World action models are zero-shot policies, 2026b. URL https://arxiv.org/abs/2602.15922.

Tianyuan Yuan, Zibin Dong, Yicheng Liu, and Hang Zhao. Fast-WAM: Do world action models need test-time future imagination?, 2026. URL https://arxiv.org/abs/2603.16666.

Jiahui Zhang, Ze Huang, Chun Gu, Zipei Ma, and Li Zhang. Reinforcing action policies by prophesying. *arXiv preprint arXiv:2511.20633*, 2025. URL https://arxiv.org/abs/2511.20633.

Qi Zhang, Shaopeng Zhai, Shengzhe Zhang, Litao Liu, Tianyi Zhang, Fuxian Huang, and Ming Zhou. A generalist pair-wise progress critic model for vision-language-action robots. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306, 2026a. URL https://openreview.net/pdf/96f0e112174f9d38d8bd80714632c2bf01c9d730.pdf.

Zongzheng Zhang, Jingrui Pang, Zhuo Yang, Kun Li, Minwen Liao, Saining Zhang, Guoxuan Chi, Jinbang Guo, Huan-ang Gao, Modi Shi, Dongyun Ge, Yao Mu, Jiayuan Gu, Rui Chen, Hao Dong, Huazhe Xu, Li Yi, Yixin Zhu, Hang Zhao, Pengwei Wang, Shanghang Zhang, Guocai Yao, Jianyu Chen, Hongyang Li, and Hao Zhao. Dexora: Open-Source VLA for High-DoF Bimanual Dexterity. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*, 2026b. URL https://dexoravla.github.io/.

Kaiwen Zheng, Huayu Chen, Haotian Ye, Haoxiang Wang, Qinsheng Zhang, Kai Jiang, Hang Su, Stefano Ermon, Jun Zhu, and Ming-Yu Liu. DiffusionNFT: Online diffusion reinforcement with forward process. In *International Conference on Learning Representations*, 2026a. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/d8e68ddfe22520b45b8fb8d5cbde5a21-Abstract-Conference.html. Oral.

Ruijie Zheng, Dantong Niu, Yuqi Xie, Jing Wang, Mengda Xu, Yunfan Jiang, Fernando Castaneda, Fengyuan Hu, You Liang Tan, Letian Fu, Trevor Darrell, Furong Huang, Yuke Zhu, Danfei Xu, and Linxi Fan. EgoScale: Scaling dexterous manipulation with diverse egocentric human data, 2026b. URL https://arxiv.org/abs/2602.16710.

Yifan Zhong, Zhang Chen, Tianrui Guan, Fanlian Zeng, Yuyao Ye, Tianjia He, Ka Nam Lui, Jiayi Li, Tingrui Zhang, Ruilin Yan, Xinhao Ji, Guangyu Zhao, Wenjie Lou, Jiayuan Zhang, Yuanpei Chen,

and Yaodong Yang. Egosteer: A full-stack system towards steerable dexterous manipulation from egocentric videos. *arXiv preprint arXiv:2607.09701*, 2026. URL https://arxiv.org/abs/2607.09701.

Fangqi Zhu, Zhengyang Yan, Zicong Hong, Quanxin Shou, Xiao Ma, and Song Guo. WMPO: World model-based policy optimization for vision-language-action models. In *International Conference on Learning Representations*, 2026a. URL https://iclr.cc/virtual/2026/poster/10007263.

Jun Zhu, Hengkai Tan, Jintao Zhang, Min Zhao, Fan Bao, and Bo Zhang. General world models from first-principles, 2026b. Manuscript.