WHAT DATA MAKES WORLD MODELS BETTER MANIPULATION PLANNERS? BEYOND SUCCESSFUL DEMONSTRATIONS WITH DIVERSE ACTION EXPERIENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

World models equip robots with the ability to predict the physical consequences of candidate actions prior to execution. However, the interaction data used to train these models is typically dominated by successful expert demonstrations. While such data ensures broad task-relevant state coverage, it offers limited experience of how alternative actions from similar states yield distinct future outcomes. In this work, we systematically investigate how the composition of robot interaction data influences world-model planning. We introduce *Diverse Action Experience* (DAE), a data-collection framework that branches alternative action trajectories from intermediate states of expert demonstrations, systematically allocating the interaction budget between broad state coverage and dense local action-outcome coverage. Evaluating multiple world-model architectures on the LIBERO benchmark, we measure performance through candidate-action ranking and closed-loop manipulation. Our experiments reveal that data composition substantially impacts planning behavior: varying state coverage, perturbation regimes, and interaction durations produces substantial differences in ranking fidelity and downstream task success. Furthermore, the preferred data composition depends strongly on the underlying world model. We also show that high aggregate ranking accuracy can obscure degenerate candidate selection, and that stronger candidate ranking does not automatically guarantee robust closed-loop control. These findings establish interaction-data composition as an important design dimension in world-model planning, alongside model architecture and planner design.

1 INTRODUCTION

World models enable robots to predict how the physical environment evolves in response to their actions. This predictive capability has long supported model-based control, allowing robots to anticipate the outcomes of candidate actions before execution. Early methods combined action-conditioned visual prediction with model predictive control for robotic manipulation (Levine & Finn, 2017). More recently, learned visual representations have extended this paradigm substantially. For example, DINO-WM learns action-conditioned dynamics in a pretrained visual feature space for test-time planning (Zhou et al., 2025), while V-JEPA 2-AC demonstrates that latent world models can support goal-conditioned planning on physical robots with limited interaction data (Assran et al., 2025). Together, these advances establish world-model planning as an important paradigm for robot decision-making: predicting possible futures, evaluating their outcomes, and executing promising actions.

Recent work has advanced world-model planning through improved visual representations, predictive architectures, training objectives, and planning algorithms. For example, recent studies systematically analyze how architectural and algorithmic choices affect physical planning performance (Terver et al., 2026), while others improve long-horizon planning through better action proposals (Cheng et al., 2026) or iterative search over model rollouts (Zhang & Du, 2026). However, these works largely treat the interaction trajectories used to train the world model as a fixed resource. A

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.

fundamental question therefore remains underexplored: *what composition of robot interaction data produces a world model that is most useful for planning?*

This question matters because planning places different demands on training data than simply reproducing expert behavior. A planner must reliably distinguish how alternative actions affect future states. Successful demonstrations provide task-relevant experience and broad state coverage, but each visited state typically contains only a narrow range of expert actions. As a result, a dataset may cover many states while providing little information about how different actions from similar states lead to different future outcomes. For robust planning, understanding the consequences of alternative actions is essential. A world model may achieve low average prediction error yet still make poor decisions if it misrepresents the relative outcomes of candidate actions. This gap motivates evaluating world models not only through prediction accuracy, but also through candidate-action ranking and closed-loop decision making (Yu et al., 2026).

In this work, we investigate how robot interaction-data composition affects world-model planning. We focus on a basic data-allocation question under a constrained interaction budget: is it more useful to cover a wider variety of task-relevant states, or to observe more diverse action outcomes around fewer states? To study this tradeoff, we construct task-relevant *Diverse Action Experience* by branching multiple alternative action trajectories from selected intermediate states of expert demonstrations. By adjusting how a fixed data-collection budget is distributed across root states and local action branches, we systematically vary broad state coverage and local action-outcome coverage. Within each controlled comparison, the model architecture and training procedure are kept fixed. At inference time, the world model predicts the consequences of candidate action sequences, an external goal scorer evaluates the predicted futures, and a receding-horizon planner executes the selected action. This controlled framework allows us to study how training data affects candidate-action ranking and downstream planning performance.

Our core contributions are threefold:

- **Data Composition for World-Model Planning.** To our knowledge, we present the first systematic study of how robot interaction-data composition affects the planning capabilities of learned world models. We study training data as a central design dimension in world-model planning, alongside model architecture and planner design.
- **Controlled Construction of Diverse Action Experience.** We introduce a structured data-generation framework that allocates a fixed interaction budget between broad state coverage and local action-outcome coverage. This enables controlled comparisons between different experience compositions under matched acquisition budgets.
- **Planning-Oriented Evaluation and Analysis.** We evaluate world models through candidate-action ranking and closed-loop planning. This evaluation identifies how the effects of data composition emerge in planning and whether they translate into downstream task performance.

2 Related Work

World Models for Robot Planning. Learned dynamics models have a long history in robot planning. Early visual model-based control combined action-conditioned prediction with model predictive control (Levine & Finn, 2017). More recent approaches perform prediction and planning in learned representation spaces. DINO-WM learns action-conditioned latent dynamics for planning toward visual goals (Zhou et al., 2025), while V-JEPA 2-AC extends latent planning to physical robot manipulation (Assran et al., 2025). End-to-end joint-embedding models such as LeWM further learn predictive representations directly from pixels (Maes et al., 2026). Subsequent studies analyze model architectures, training objectives, and planning algorithms (Terver et al., 2026), as well as action proposals and iterative plan refinement (Cheng et al., 2026; Zhang & Du, 2026). Our work studies a complementary dimension: the composition of the interaction data used to train the world model.

Data Composition and Diversity in Robot Learning. Large-scale datasets such as Open X-Embodiment (Collaboration et al., 2025) and DROID (Khazatsky et al., 2024) demonstrate that diversity across tasks, environments, and robot embodiments can improve robot learning and gen-

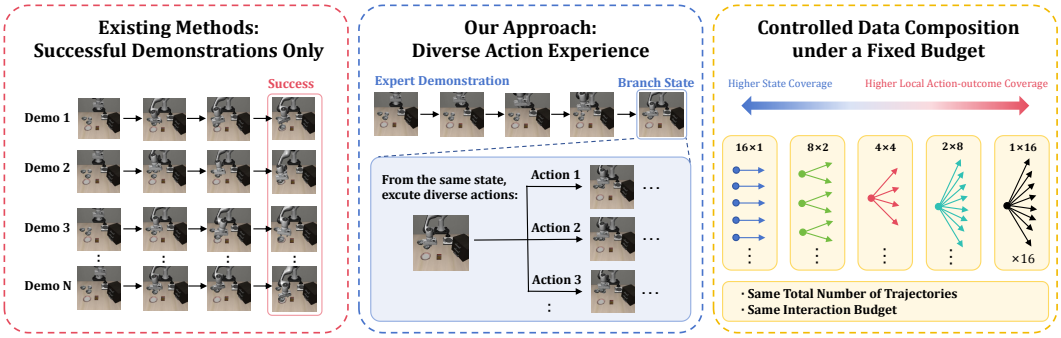


Figure 1: **Overview of our data-centric study of world-model planning.** (a) Successful demonstrations provide broad task-relevant state coverage but limited local action variation. (b) We construct Diverse Action Experience (DAE) by branching multiple action trajectories from selected intermediate states of expert demonstrations, collecting different action outcomes around the same state. (c) Under a fixed interaction budget, we vary the allocation between the number of root states and local action branches, enabling a controlled study of state coverage versus local action-outcome coverage.

eralization. More targeted studies show that data composition also matters. Data-scaling analyses indicate that environment and object diversity can provide greater benefits than repeatedly collecting demonstrations in the same settings (Lin et al., 2025), while other work studies how different forms of diversity affect policy learning (Shi et al., 2025). Frameworks such as MimicGen further show that diverse successful trajectories can be generated from a small set of human demonstrations (Mandlekar et al., 2023). These studies primarily focus on data composition for model-free policy learning. In contrast, we study data composition for action-conditioned world models, with a focus on the tradeoff between broad state coverage and local action-outcome coverage for planning.

Evaluating World Models for Decision Making. World models are commonly evaluated using reconstruction error, feature prediction error, or generative metrics. While these metrics measure prediction quality, they do not necessarily indicate usefulness for downstream planning. Recent work therefore emphasizes decision-oriented evaluation, including whether predicted outcomes remain accurate under alternative actions, whether models preserve action or policy rankings, and whether they improve closed-loop performance (Yu et al., 2026). Physical planning studies similarly use downstream task success as an important measure of world-model utility (Terver et al., 2026). Following this perspective, we jointly evaluate candidate-action ranking and closed-loop task success to study how training-data composition affects planning-relevant world-model capabilities.

3 Method

We study world-model planning from a data-centric perspective. Rather than introducing a new architecture or planning algorithm, we systematically vary the interaction experience used for training and measure how data composition affects planning performance. Our framework consists of four main stages: defining data composition, constructing Diverse Action Experience (DAE) from task-relevant states, varying state and local action-outcome coverage under a fixed budget, and evaluating world-model planning across these data conditions. Figure 1 provides an overview.

3.1 Data Composition for World-Model Planning

We consider an offline robot dataset composed of trajectories

$$\tau = \{(o_t, p_t, a_t)\}_{t=1}^{T},$$

where o_t is the visual observation, p_t is robot proprioception, and a_t is the executed action. An action-conditioned world model is trained to predict how the observed state evolves under a sequence of actions. At test time, a planner uses the model to evaluate and compare the predicted outcomes of alternative action sequences.

This setting places different demands on data than standard behavioral cloning. A successful demonstration provides a sequence of task-relevant states paired with expert actions. Collecting demonstrations from diverse initial conditions can provide broad *state coverage*, but each visited state usually contains limited action variation. Consequently, the model sees little evidence of how alternative actions from similar states lead to different future outcomes.

We therefore define two properties of interaction data. **State coverage** measures the breadth of task-relevant robot and scene states represented in the dataset. **Local action-outcome coverage** measures the diversity of physical outcomes observed when different actions are executed from similar states. Our central question is how these two forms of coverage should be balanced under a limited robot interaction budget.

3.2 Diverse Action Experience

To increase local action-outcome coverage while maintaining task relevance, we introduce *Diverse Action Experience* (DAE). The central idea is to branch multiple alternative action trajectories from intermediate states of successful demonstrations.

Task-relevant root states. For each selected expert demonstration, we replay a short expert prefix to reach an intermediate state along a valid task trajectory, which we call a *root state*. Using demonstration states as roots keeps data collection focused on task-relevant regions of the state space. The execution cost of the expert prefix is included in the physical interaction budget.

Local action branching. From each root state, we generate multiple task-relevant action branches. Each branch follows the overall structure of the manipulation task while varying local execution factors such as target approach, grasp location, transport path, and release location. This produces different action sequences and physical outcomes from nearly identical starting states. During data collection, the controller may use privileged simulator information, such as exact object and end-effector poses, to construct these branches. This privileged information is not provided to the world model during training or planning.

Outcome-preserving collection. DAE is designed to expose the world model to diverse physical outcomes rather than assign quality labels to trajectories. We therefore do not require every branch to succeed, nor do we provide success, reward, or progress labels to the model. Branches are retained according to the collection protocol rather than final task success, allowing the model to observe the physical outcomes of the executed actions. Only the newly collected branch suffixes are added as training experience; the expert prefix is counted only toward the acquisition cost.

3.3 Controlled Data Composition

We use DAE to construct controlled datasets that vary how a fixed interaction budget is allocated. Let N_r denote the number of distinct root states and N_b the number of branches collected per root. For the root–branch allocation study, we fix the total number of newly collected trajectories:

$$N_r N_b = N_{\mathrm{traj}}.$$

Increasing N_r broadens state coverage but reduces the number of alternative actions observed around each root. Increasing N_b instead concentrates the budget on denser local action-outcome coverage. We study a range of allocations, from many roots with few branches per root to fewer roots with many branches per root.

All data conditions share a base dataset of successful demonstrations, $\mathcal{D}_0$. The world-model training set for condition c is

$$\mathcal{D}^{(c)} = \mathcal{D}_0 \cup \mathcal{D}_{\mathrm{DAE}}^{(c)}.$$

Within each world-model family, we keep the physical interaction budget, model architecture, and planning protocol fixed across the corresponding controlled comparison. The primary experimental variable is the distribution of interaction experience across root states and local action branches.

We distinguish the *interaction budget*, measured by environment steps used for data acquisition, from *optimization exposure*, which describes how often each data source is sampled during training.

3.4 Learning from Different Experience Compositions

Our primary controlled study follows the latent predictive formulation of DINO-WM (Zhou et al., 2025). Within each world-model family, the architecture and learning objective are kept fixed across data conditions. DINO-WM receives multi-view RGB observations, robot proprioception, and actions. A frozen visual encoder extracts spatial features, while learned encoders map proprioception and actions into the predictive representation. A predictive model then learns the action-conditioned dynamics of these representations.

Training samples are extracted as short causal windows of observations and actions. At each timestep, the model predicts the representation of the next observation. The DINO-WM objective combines visual and proprioceptive prediction losses:

$$\mathcal{L} = \mathcal{L}_{\text{vis}} + \lambda_p \mathcal{L}_{\text{prop}},$$

where λ_p weights the proprioceptive prediction term.

The model is trained only through predictive learning and receives no task-success, reward, or progress labels. Within each model family, demonstrations and DAE trajectories use the same training objective and differ in the state-action-outcome distributions they contain. We additionally evaluate the world-model formulations introduced in Sec. 4.1 under selected data-composition settings, using the prediction objective native to each model.

3.5 Planning with the Learned World Model

We evaluate each trained model using a standardized closed-loop planning procedure, illustrated in Fig. 2. At each control step t, a shared proposal mechanism retrieves a set of candidate action sequences:

$$\mathcal{A}_t = \{A_t^{(1)}, A_t^{(2)}, \ldots, A_t^{(K)}\}.$$

The same proposal mechanism is used across compared models and data conditions, keeping candidate generation as a common component of the planning system.

For each candidate $A_t^{(i)}$, the world model predicts its future state in the model's predictive representation:

$$\hat{z}_{t+H}^{(i)} = f_\theta\left(h_t, A_t^{(i)}\right),$$

where h_t denotes the model-specific representation of the current observation history.

An external goal scorer assigns a goal cost to each predicted future:

$$J^{(i)} = d\left(\hat{z}_{t+H}^{(i)}, z_g\right),$$

where z_g is the encoded goal and d is a feature-space distance metric. The planner selects the candidate with the lowest predicted cost.

We use receding-horizon control. The robot executes a short segment of the selected sequence, observes the resulting state, retrieves a new candidate set, and replans. This predict–select–execute–replan cycle continues until task completion or the evaluation horizon is reached.

In this setup, the world model acts as a predictive evaluator rather than a direct policy. The proposal mechanism and goal scorer are shared across comparisons. In Sec. 4, we evaluate candidate-action ranking and closed-loop planning performance across different data compositions.

4 Experiments

In this section, we investigate four primary questions: **Q1:** How does interaction-data composition affect world-model planning? **Q2:** Which properties of Diverse Action Experience (DAE) yield the most benefit? **Q3:** How accurately do world models rank candidate actions? **Q4:** Do differences in candidate ranking translate into successful closed-loop manipulation? We address Q1–Q3 in the main text and present the detailed closed-loop analysis for Q4 in Appendix B.

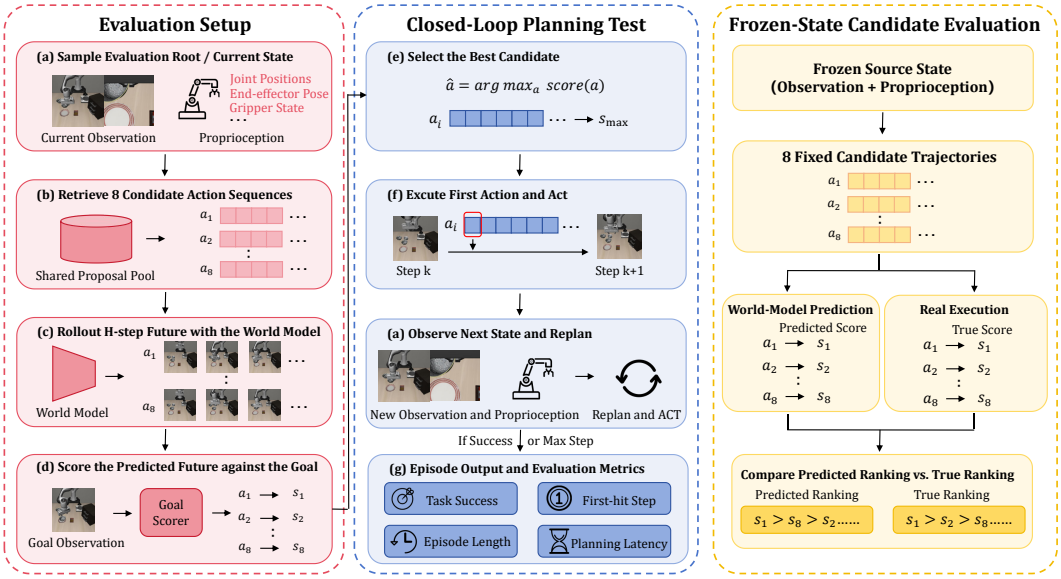


Figure 2: **Evaluation protocol for world-model planning. Left:** Eight candidate action sequences are retrieved from a shared proposal pool and evaluated using world-model predictions. **Middle:** The lowest-cost candidate is selected and executed, after which the robot observes the new state and replans. **Right:** For candidate evaluation, the same candidates are predicted and physically executed from the same source state, enabling direct comparison between predicted and realized rankings.

4.1 Experimental Setup

Benchmark and data. We conduct our evaluations on the LIBERO benchmark (Liu et al., 2023). All data conditions are built upon a shared base dataset of successful demonstrations $\mathcal{D}_0$, varying exclusively in the supplementary task-relevant interaction experience detailed in Sec. 3.2. We establish D-EXP as the reference DAE condition, constructing D-BS and D-LS to emphasize broader state coverage and denser local action-outcome coverage, respectively. Furthermore, we systematically ablate the action perturbation regime and the temporal duration of the collected segments.

World models. We evaluate four distinct world-model architectures, each employing a unique predictive representation and learning objective: DINO-WM (Zhou et al., 2025), which models dynamics over frozen DINOv2 features; LeWM (Maes et al., 2026), an end-to-end joint-embedding predictive model; JEPA-WM (Terver et al., 2026), a joint-embedding architecture tailored for physical planning; and a standard reconstruction-based predictive model.

Evaluation. Candidate-action ranking is assessed from shared source states using $K = 8$ candidate action sequences. Closed-loop control is evaluated against the official LIBERO task predicates. We report *Any Success* (whether the task predicate is satisfied at any juncture during the episode) and *Terminal Success* (whether the predicate holds at the final timestep). To ensure parity, all paired comparisons utilize identical evaluation roots.

4.2 Evaluation Protocol

Figure 2 illustrates our unified planning and evaluation pipeline. Given the current visual observation and robot proprioception, a shared proposal mechanism yields $K = 8$ candidate action sequences. The world model predicts the future consequence of each candidate, which is subsequently scored by an external goal function.

For candidate-action ranking, all models process an identical source state s_t and candidate set $\{A_t^{(i)}\}_{i=1}^{K}$. The predicted cost of candidate i is formulated as

$$J_{\text{pred}}^{(i)} = d\left(\hat{z}_{t+H}^{(i)}, z_g\right),$$

Table 1: **Effect of interaction-data composition on world-model planning.** All metrics denote percentages. Top-1 accuracy is measured at $H = 10$, while Any Success reflects closed-loop task completion.

Data Factor	World Model	Metric	Condition 1	Condition 2	Condition 3
State / Local Coverage	JEPA-WM	Top-1	$\mathcal{D}_0$: 56.0	D-BS: 26.0	D-LS: 50.0
State / Local Coverage	Reconstruction WM	Top-1	$\mathcal{D}_0$: 64.0	D-BS: 64.0	D-LS: 64.0
Perturbation	DINO-WM	Any Success	Small/Short: 20.0	Reference: 10.0	Large/Long: 17.5
Perturbation	LeWM	Any Success	Small/Short: 32.5	Reference: 35.0	Large/Long: 0.0
Duration	DINO-WM	Any Success	Short: 10.0	Mixed: 25.0	Long: 15.0
Duration	LeWM	Any Success	Short: 50.0	Mixed: 1.7	Long: 35.0

while deploying the corresponding candidate physically in the environment yields

$$J_{\text{true}}^{(i)} = d\left(z_{t+H}^{(i)}, z_g\right).$$

If the world model ultimately selects

$$\hat{i} = \arg\min_i J_{\text{pred}}^{(i)},$$

the incurred ranking regret is defined as

$$\text{Regret} = J_{\text{true}}^{(\hat{i})} - \min_i J_{\text{true}}^{(i)}.$$

We report Top-1 agreement and regret for candidate-action ranking, and the official task success rates for closed-loop manipulation.

4.3 Q1: How Does Data Composition Affect World-Model Planning?

The structural composition of interaction data yields profound differences in both candidate-action ranking and closed-loop planning. Table 1 summarizes the performance shifts induced by varying state coverage, local action outcomes, perturbation regime, and segment duration.

This impact is consistent across multiple data-allocation factors. For instance, when modifying state and local coverage, JEPA-WM's performance under D-BS drops to 26.0% Top-1 accuracy at $H = 10$, compared to 56.0% for $\mathcal{D}_0$ and 50.0% for D-LS. Conversely, the reconstruction-based baseline stagnates at 64.0% regardless of the data condition.

Closed-loop task success is similarly sensitive to DAE composition. Changing the perturbation regime reduces LeWM's success rate from 35.0% under the reference condition to 0.0% under the large/long condition. Variations in interaction duration produce similarly large differences, with LeWM ranging from 1.7% to 50.0% and DINO-WM from 10.0% to 25.0%. These findings show that the structure of interaction data is an important determinant of planning performance, and that the preferred data composition differs across world models.

4.4 Q2: Which Properties of Diverse Action Experience Matter?

The efficacy of DAE relies heavily on both the perturbation regime and the duration of the collected segments. We examine these two properties in Table 2.

Across perturbation regimes, DINO-WM exhibits relative robustness between the small/short and large/long conditions, with 20.0% and 17.5% Any Success. In contrast, LeWM is more sensitive to this factor; its Any Success decreases from 32.5% to 0.0%, representing a difference of 32.5 percentage points with a 95% paired bootstrap CI of $[22.5, 42.5]$.

Interaction duration exposes a distinct dynamic. DINO-WM peaks under mixed-duration experience, achieving 25.0% Any Success, compared with 10.0% for short and 15.0% for long segments. The improvement from short to mixed is 15.0 percentage points with a 95% CI of $[1.7, 26.7]$. LeWM, however, performs best with short interaction segments at 50.0%, reaches 1.7% under the mixed condition, and reaches 35.0% with long segments. Consequently, there is no single perturbation regime or interaction duration that performs best across both world models.

Table 2: **Effect of perturbation regime and interaction duration.** We report Any Success and Terminal Success for DINO-WM and LeWM.

Factor	World Model	Condition	Any Success ↑	Terminal Success ↑
Perturbation	DINO-WM	Small / Short	20.0%	20.0%
	DINO-WM	Reference	10.0%	5.0%
	DINO-WM	Large / Long	17.5%	7.5%
	LeWM	Small / Short	32.5%	30.0%
	LeWM	Reference	35.0%	30.0%
	LeWM	Large / Long	0.0%	0.0%
Duration	DINO-WM	Short	10.0%	8.3%
	DINO-WM	Mixed	25.0%	18.3%
	DINO-WM	Long	15.0%	10.0%
	LeWM	Short	50.0%	41.7%
	LeWM	Mixed	1.7%	1.7%
	LeWM	Long	35.0%	28.3%

Table 3: **Long-horizon candidate ranking.** Entries signify Top-1 candidate hits evaluated over 50 held-out source states.

World Model	Data	$H = 1$	$H = 5$	$H = 10$
DINO-WM	$\mathcal{D}_0$	28	30	24
LeWM	$\mathcal{D}_0$	28	25	21
JEPA-WM	$\mathcal{D}_0$	30	27	28
JEPA-WM	D-BS	32	28	13
JEPA-WM	D-LS	33	27	25
Reconstruction WM	$\mathcal{D}_0$	32	32	32
Reconstruction WM	D-BS	32	32	32
Reconstruction WM	D-LS	31	33	32

4.5 Q3: How Well Do World Models Rank Candidate Actions?

Evaluating candidate-action ranking not only highlights distinct capabilities among world models but also exposes critical failure modes obscured by aggregate accuracy metrics.

During short-horizon evaluation, we execute 3,600 candidate trajectories physically from shared source states. Across nine evaluated model–data–horizon combinations, DINO-WM achieves better candidate-ranking performance than LeWM in eight. Since these candidates are benchmarked against actual physical outcomes, this protocol serves as a direct diagnostic of whether a model's internal predicted ordering aligns with physical reality.

To push this further, we assess long-horizon ranking across 50 held-out source states (eight candidates per state), with each sequence executed for 180 control steps.

Ranking fidelity varies strongly across models and datasets. At $H = 10$, JEPA-WM obtains 28/50 Top-1 hits using $\mathcal{D}_0$, decreases to 13/50 with D-BS, and reaches 25/50 with D-LS. Under $\mathcal{D}_0$, DINO-WM obtains 24/50 Top-1 hits at $H = 10$, compared with 21/50 for LeWM.

Notably, the reconstruction-based model nominally appears competitive, recording 32/50 hits at $H = 10$. However, closer inspection reveals degenerate behavior: under $\mathcal{D}_0$, it deterministically selects candidate slot 0 for all 50 states. A fixed-slot-0 baseline exactly matches this 32/50 score, while an oracle selection reaches 40/50. The model's high Top-1 accuracy therefore comes from a fixed candidate-selection pattern rather than state-dependent candidate discrimination.

5 Conclusion

In this work, we investigated world-model planning through the lens of interaction-data composition. Moving beyond the paradigm of treating robot experience as a static training resource, we system-

atically analyzed how different allocations of task-relevant interaction data affect candidate-action ranking and closed-loop manipulation. To facilitate this, we introduced Diverse Action Experience (DAE), a framework that branches alternative action trajectories from intermediate states of successful demonstrations. DAE explicitly enhances local action-outcome coverage without relying on task success, reward, or progress labels.

Our empirical results demonstrate that the composition of interaction experience substantially affects world-model planning capabilities. Shifting the balance between broad state coverage and local action-outcome coverage alters long-horizon candidate ranking, while varying perturbation regimes and interaction durations produces large changes in closed-loop task success. These effects also differ across world models, showing that there is no single data configuration that performs best across all models.

Furthermore, our analysis highlights the limitations of relying solely on aggregate planning metrics for model evaluation. We showed that a reconstruction-based model can achieve competitive Top-1 ranking scores while exhibiting degenerate candidate selection. Our closed-loop evaluations further show that candidate ranking alone is insufficient for successful manipulation, as candidate generation and action selection also affect planning performance.

Ultimately, our findings establish interaction-data composition as a foundational design parameter for learned world models. Expanding beyond the aggregation of successful demonstrations, actively exposing a model to diverse local action outcomes provides a direct mechanism to shape its learned planning capabilities. We hope this data-centric perspective encourages future research to study robot interaction data and world-model design together rather than treating them as separate components.

AI Use Statement

Generative AI tools were used to assist with language polishing and improving the readability of the manuscript. They were also used to assist with implementing and debugging research code based on methods and designs specified by the authors. All AI-assisted code was manually reviewed, tested, and validated by the authors. The authors reviewed all AI-assisted manuscript edits and take full responsibility for the final content, experimental results, and conclusions of this work.

References

Mahmoud Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Mojtaba Komeili, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, Sergio Arnaud, Abha Gejji, Ada Martin, Francois Robert Hogan, Daniel Dugas, Piotr Bojanowski, Vasil Khalidov, Patrick Labatut, Francisco Massa, Marc Szafraniec, Kapil Krishnakumar, Yong Li, Xiaodong Ma, Sarath Chandar, Franziska Meier, Yann LeCun, Michael Rabbat, and Nicolas Ballas. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025.

Letian Cheng, Qi Zhang, and Yisen Wang. Sage: Subgoal-conditioned action generation for latent world model planning, 2026. URL https://arxiv.org/abs/2607.17973.

Embodiment Collaboration, Abby O'Neill, Abdul Rehman, Abhinav Gupta, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, Albert Tung, Alex Bewley, Alex Herzog, Alex Irpan, Alexander Khazatsky, Anant Rai, Anchit Gupta, Andrew Wang, Andrey Kolobov, Anikait Singh, Animesh Garg, Aniruddha Kembhavi, Annie Xie, Anthony Brohan, Antonin Raffin, Archit Sharma, Arefeh Yavary, Arhan Jain, Ashwin Balakrishna, Ayzaan Wahid, Ben Burgess-Limerick, Beomjoon Kim, Bernhard Schölkopf, Blake Wulfe, Brian Ichter, Cewu Lu, Charles Xu, Charlotte Le, Chelsea Finn, Chen Wang, Chenfeng Xu, Cheng Chi, Chenguang Huang, Christine Chan, Christopher Agia, Chuer Pan, Chuyuan Fu, Coline Devin, Danfei Xu, Daniel Morton, Danny Driess, Daphne Chen, Deepak Pathak, Dhruv Shah, Dieter Büchler, Dinesh Jayaraman, Dmitry Kalashnikov, Dorsa Sadigh, Edward Johns, Ethan Foster, Fangchen Liu, Federico Ceola, Fei Xia, Feiyu Zhao, Felipe Vieira Frujeri, Freek Stulp, Gaoyue Zhou, Gaurav S. Sukhatme, Gautam Salhotra, Ge Yan, Gilbert Feng, Giulio Schiavi, Glen Berseth, Gregory Kahn, Guangwen Yang, Guanzhi Wang, Hao Su, Hao-Shu Fang, Haochen Shi, Henghui Bao, Heni Ben Amor, Henrik I Christensen,

Hiroki Furuta, Homanga Bharadhwaj, Homer Walke, Hongjie Fang, Huy Ha, Igor Mordatch, Ilija Radosavovic, Isabel Leal, Jacky Liang, Jad Abou-Chakra, Jaehyung Kim, Jaimyn Drake, Jan Peters, Jan Schneider, Jasmine Hsu, Jay Vakil, Jeannette Bohg, Jeffrey Bingham, Jeffrey Wu, Jensen Gao, Jiaheng Hu, Jiajun Wu, Jialin Wu, Jiankai Sun, Jianlan Luo, Jiayuan Gu, Jie Tan, Jihoon Oh, Jimmy Wu, Jingpei Lu, Jingyun Yang, Jitendra Malik, João Silvério, Joey Hejna, Jonathan Booher, Jonathan Tompson, Jonathan Yang, Jordi Salvador, Joseph J. Lim, Junhyek Han, Kaiyuan Wang, Kanishka Rao, Karl Pertsch, Karol Hausman, Keegan Go, Keerthana Gopalakrishnan, Ken Goldberg, Kendra Byrne, Kenneth Oslund, Kento Kawaharazuka, Kevin Black, Kevin Lin, Kevin Zhang, Kiana Ehsani, Kiran Lekkala, Kirsty Ellis, Krishan Rana, Krishnan Srinivasan, Kuan Fang, Kunal Pratap Singh, Kuo-Hao Zeng, Kyle Hatch, Kyle Hsu, Laurent Itti, Lawrence Yunliang Chen, Lerrel Pinto, Li Fei-Fei, Liam Tan, Linxi "Jim" Fan, Lionel Ott, Lisa Lee, Luca Weihs, Magnum Chen, Marion Lepert, Marius Memmel, Masayoshi Tomizuka, Masha Itkina, Mateo Guaman Castro, Max Spero, Maximilian Du, Michael Ahn, Michael C. Yip, Mingtong Zhang, Mingyu Ding, Minho Heo, Mohan Kumar Srirama, Mohit Sharma, Moo Jin Kim, Muhammad Zubair Irshad, Naoaki Kanazawa, Nicklas Hansen, Nicolas Heess, Nikhil J Joshi, Niko Suenderhauf, Ning Liu, Norman Di Palo, Nur Muhammad Mahi Shafiullah, Oier Mees, Oliver Kroemer, Osbert Bastani, Pannag R Sanketi, Patrick "Tree" Miller, Patrick Yin, Paul Wohlhart, Peng Xu, Peter David Fagan, Peter Mitrano, Pierre Sermanet, Pieter Abbeel, Priya Sundaresan, Qiuyu Chen, Quan Vuong, Rafael Rafailov, Ran Tian, Ria Doshi, Roberto Martín-Martín, Rohan Baijal, Rosario Scalise, Rose Hendrix, Roy Lin, Runjia Qian, Ruohan Zhang, Russell Mendonca, Rutav Shah, Ryan Hoque, Ryan Julian, Samuel Bustamante, Sean Kirmani, Sergey Levine, Shan Lin, Sherry Moore, Shikhar Bahl, Shivin Dass, Shubham Sonawani, Shubham Tulsiani, Shuran Song, Sichun Xu, Siddhant Haldar, Siddharth Karamcheti, Simeon Adebola, Simon Guist, Soroush Nasiriany, Stefan Schaal, Stefan Welker, Stephen Tian, Subramanian Ramamoorthy, Sudeep Dasari, Suneel Belkhale, Sungjae Park, Suraj Nair, Suvir Mirchandani, Takayuki Osa, Tanmay Gupta, Tatsuya Harada, Tatsuya Matsushima, Ted Xiao, Thomas Kollar, Tianhe Yu, Tianli Ding, Todor Davchev, Tony Z. Zhao, Travis Armstrong, Trevor Darrell, Trinity Chung, Vidhi Jain, Vikash Kumar, Vincent Vanhoucke, Vitor Guizilini, Wei Zhan, Wenxuan Zhou, Wolfram Burgard, Xi Chen, Xiangyu Chen, Xiaolong Wang, Xinghao Zhu, Xinyang Geng, Xiyuan Liu, Xu Liangwei, Xuanlin Li, Yansong Pang, Yao Lu, Yecheng Jason Ma, Yejin Kim, Yevgen Chebotar, Yifan Zhou, Yifeng Zhu, Yilin Wu, Ying Xu, Yixuan Wang, Yonatan Bisk, Yongqiang Dou, Yoonyoung Cho, Youngwoon Lee, Yuchen Cui, Yue Cao, Yueh-Hua Wu, Yujin Tang, Yuke Zhu, Yunchu Zhang, Yunfan Jiang, Yunshuang Li, Yunzhu Li, Yusuke Iwasawa, Yutaka Matsuo, Zehan Ma, Zhuo Xu, Zichen Jeff Cui, Zichen Zhang, Zipeng Fu, and Zipeng Lin. Open x-embodiment: Robotic learning datasets and rt-x models, 2025. URL https://arxiv.org/abs/2310.08864.

Alexander Khazatsky, Karl Pertsch, Suraj Nair, Ashwin Balakrishna, Sudeep Dasari, Siddharth Karamcheti, Soroush Nasiriany, Mohan Kumar Srirama, Lawrence Yunliang Chen, Kirsty Ellis, Peter David Fagan, Joey Hejna, Masha Itkina, Marion Lepert, Yecheng Jason Ma, Patrick Tree Miller, Jimmy Wu, Suneel Belkhale, Shivin Dass, Huy Ha, Arhan Jain, Abraham Lee, Youngwoon Lee, Marius Memmel, Sungjae Park, Ilija Radosavovic, Kaiyuan Wang, Albert Zhan, Kevin Black, Cheng Chi, Kyle Beltran Hatch, Shan Lin, Jingpei Lu, Jean Mercat, Abdul Rehman, Pannag R Sanketi, Archit Sharma, Cody Simpson, Quan Vuong, Homer Rich Walke, Blake Wulfe, Ted Xiao, Jonathan Heewon Yang, Arefeh Yavary, Tony Z. Zhao, Christopher Agia, Rohan Baijal, Mateo Guaman Castro, Daphne Chen, Qiuyu Chen, Trinity Chung, Jaimyn Drake, Ethan Paul Foster, Jensen Gao, Vitor Guizilini, David Antonio Herrera, Minho Heo, Kyle Hsu, Jiaheng Hu, Muhammad Zubair Irshad, Donovon Jackson, Charlotte Le, Yunshuang Li, Kevin Lin, Roy Lin, Zehan Ma, Abhiram Maddukuri, Suvir Mirchandani, Daniel Morton, Tony Nguyen, Abigail O'Neill, Rosario Scalise, Derick Seale, Victor Son, Stephen Tian, Emi Tran, Andrew E. Wang, Yilin Wu, Annie Xie, Jingyun Yang, Patrick Yin, Yunchu Zhang, Osbert Bastani, Glen Berseth, Jeannette Bohg, Ken Goldberg, Abhinav Gupta, Abhishek Gupta, Dinesh Jayaraman, Joseph J Lim, Jitendra Malik, Roberto Martín-Martín, Subramanian Ramamoorthy, Dorsa Sadigh, Shuran Song, Jiajun Wu, Michael C. Yip, Yuke Zhu, Thomas Kollar, Sergey Levine, and Chelsea Finn. Droid: A large-scale in-the-wild robot manipulation dataset. 2024.

Sergey Levine and Chelsea Finn. Deep visual foresight for planning robot motion. *ICRA*, 2017. URL https://arxiv.org/abs/1610.00696.

Fanqi Lin, Yingdong Hu, Pingyue Sheng, Chuan Wen, Jiacheng You, and Yang Gao. Data scaling laws in imitation learning for robotic manipulation. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 54877–54910, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/88b7b2c896506daabc8d3fd587055167-Paper-Conference.pdf.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. Libero: Benchmarking knowledge transfer for lifelong robot learning, 2023. URL https://arxiv.org/abs/2306.03310.

Lucas Maes, Quentin Le Lidec, Damien Scieur, Yann LeCun, and Randall Balestriero. Leworld-model: Stable end-to-end joint-embedding predictive architecture from pixels, 2026. URL https://arxiv.org/abs/2603.19312.

Ajay Mandlekar, Soroush Nasiriany, Bowen Wen, Iretiayo Akinola, Yashraj Narang, Linxi Fan, Yuke Zhu, and Dieter Fox. Mimicgen: A data generation system for scalable robot learning using human demonstrations. In Jie Tan, Marc Toussaint, and Kourosh Darvish (eds.), *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pp. 1820–1864. PMLR, 06–09 Nov 2023. URL https://proceedings.mlr.press/v229/mandlekar23a.html.

Modi Shi, Li Chen, Jin Chen, Yuxiang Lu, Chiming Liu, Guanghui Ren, Ping Luo, Di Huang, Maoqing Yao, and Hongyang Li. Is diversity all you need for scalable robotic manipulation? *arXiv preprint arXiv:2507.06219*, 2025.

Basile Terver, Tsung-Yen Yang, Jean Ponce, Adrien Bardes, and Yann LeCun. What drives success in physical planning with joint-embedding predictive world models? *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=cHZn5Gdh8e. Reproducibility Certification.

Yang Yu, Shiyuan Zhang, Yifei Sheng, Haoxiang Ren, and Haoxin Lin. How should world models be evaluated for embodied decision-making? a decision-making-centric position, 2026. URL https://arxiv.org/abs/2606.15032.

Xiangcheng Zhang and Yilun Du. World action planner: Generalizable decision-making with action-conditioned world models, 2026. URL https://arxiv.org/abs/2607.27599.

Gaoyue Zhou, Hengkai Pan, Yann Lecun, and Lerrel Pinto. DINO-WM: World models on pre-trained visual features enable zero-shot planning. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 79115–79135. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/zhou25t.html.

A Additional Evaluation Details

Closed-loop success. During the 180-step closed-loop evaluation, *Any Success* indicates that the official LIBERO task predicate is triggered at least once throughout the episode. Conversely, *Terminal Success* demands that the predicate remains satisfied at the terminal timestep. Identical initialization states enforce strict parity across conditions.

Fixed-candidate baseline. We benchmark against a deterministic slot-0 baseline, which unconditionally executes the first proposed candidate sequence, bypassing world-model ranking entirely. On the 20-root subset leveraged for DAE property ablation, this strategy achieves a 12/20 Any Success rate and a 12/20 Terminal Success rate.

Long-horizon candidate evaluation. To compute long-horizon ranking accuracy, we physically unroll eight proposed action sequences from each of the 50 validation states. We then contrast the specific candidate favored by the world model against the actual physical outcomes. The fixed-slot baseline reaches 32/50, while hindsight optimal selection establishes an upper performance bound of 40/50.

B Q4: Do Ranking Differences Translate to Closed-Loop Manipulation?

Candidate-ranking differences do not always translate directly into stronger closed-loop task performance. We first evaluate DINO-WM, LeWM, and the fixed-slot baseline on 50 paired evaluation states.

Table 4: **Closed-loop planning performance on 50 paired evaluation states.**

Controller	Any Success ↑	Terminal Success ↑
Fixed slot-0	54.0%	52.0%
DINO-WM	18.0%	12.0%
LeWM	8.0%	2.0%

The fixed-slot baseline achieves 54.0% Any Success, compared with 18.0% for DINO-WM and 8.0% for LeWM. This corresponds to differences of 36.0 and 46.0 percentage points, respectively. These results show that candidate ranking alone does not determine closed-loop success. Candidate generation and repeated action selection remain important parts of the planning system.

We further evaluate DINO-WM and LeWM on 200 paired initial states under a shared planning interface. DINO-WM succeeds on 122 states, whereas LeWM succeeds on 71.

Table 5: **Closed-loop planning under a shared planning interface.**

World Model	Success	Success Rate
DINO-WM	122 / 200	61.0%
LeWM	71 / 200	35.5%

DINO-WM achieves a 25.5 percentage-point higher paired success rate than LeWM, with a 95% paired bootstrap confidence interval of $[16.5, 34.5]$ percentage points. This result shows that differences observed between the two world models remain substantial during closed-loop execution.

Together, the two evaluations show that candidate ranking and closed-loop manipulation capture related but distinct aspects of world-model planning. Ranking quality affects action selection, while final task performance also depends on the candidate set and the repeated interaction between prediction, selection, execution, and replanning.

C Recovery-Oriented Data Collection

We additionally investigate whether the DAE collection protocol can generate recovery-oriented experience. Using the same 40 root states, we compare ordinary continuations against recovery-oriented branches rolled out over a 60-step horizon. A sequence is counted as returning when it reaches a predefined task-relevant state region.

Table 6: **Recovery behavior during interaction-data synthesis.**

Collection Strategy	Returned States	Return Rate
Ordinary continuation	0 / 40	0.0%
Recovery continuation	29 / 40	72.5%

Recovery-oriented continuations return to the task-relevant state region in 29 out of 40 instances, corresponding to 72.5%. Ordinary continuations do not return in any of the 40 cases. These results show that the DAE collection procedure can generate substantially more trajectories that return toward task-relevant states.