CALIBRATING IMAGINATION BY DECISION REGRET

Anonymous authors
Paper under double-blind review

ABSTRACT

Action-conditioned world models can replace some camera updates, but visual prediction error alone does not measure the consequence of acting on an imagined observation. DREAM-CAL estimates policy-conditioned decision regret from paired continuations and selects a refresh horizon using a block-calibrated upper bound. Across 37 tasks at nominal 25% visual refresh, DREAM-CAL achieves 1,575/2,220 (70.95%) versus 1,392/2,220 (62.70%) for a visual-uncertainty head, a difference of 8.24 percentage points from the integer counts. In a separate segment audit, harmful-event AUROC is 0.790 versus 0.665, while the visual head retains lower expected calibration error. Gains vary by task family, and four of 37 task counts favor the visual head. We evaluate closed-loop success, event ranking, regret-bound coverage, distribution shift, and inference cost. Under dynamics shift, coverage with the original calibration offset falls from 87.92% to 71.25%, and mean decision computation takes 452 ms.

1 INTRODUCTION

A robot approaching a socket can continue without a new camera image while its pose estimate remains adequate for alignment. Near contact, the same interval without feedback may allow a small pose error to turn insertion into jamming. An action-conditioned world model can fill the missing observation slots, but image fidelity alone does not identify when this substitution is useful. A changed background may alter many predicted pixels while leaving the policy's action nearly unchanged. An error localized to the contact geometry may affect fewer pixels and change the action that determines success. Figure 1 illustrates both cases and a third, in which several plausible futures lead to actions in the same region.

This distinction matters when visual updates have a cost. Action-chunking policies predict several future actions at once (Chi et al., 2023; Zhao et al., 2023), and the value of additional sensing varies with the task and the consuming policy (Krishna et al., 2025). Prediction uncertainty offers one way to schedule feedback. Recent methods estimate dense confidence in generated video (Mei et al., 2025) or structure uncertainty across prediction horizons (Wan et al., 2026). These quantities describe the prediction process. A refresh decision also depends on how the current policy responds to that prediction and on the consequence of executing its response.

We use a paired continuation to define that consequence. From a restored state, one branch gives the frozen policy visual feedback at every step. A second branch replaces those observations with world-model predictions for a candidate horizon, then returns to full feedback. Their difference in task return, together with avoidable failure and contact cost, defines a cost-augmented decision regret. The target is specific to a policy and an execution horizon: a future can be inaccurate yet harmless for one action, or plausible yet harmful for another. State restoration is needed to collect supervision, while the deployed score uses only information available before the next image is acquired.

DREAM-CAL estimates this target with policy-conditioned features. It samples predicted futures, evaluates the frozen policy on them with shared noise for stochastic policies, and combines action disagreement with the proposed chunk, proprioception and visual uncertainty. A learned head predicts regret quantiles and the probability of a harmful segment. A trajectory-block conformal offset adjusts the upper quantile, and the controller selects an admissible interval without new visual feedback.

Our evaluation separates four quantities that are often conflated: ranking harmful segments, calibrating their probabilities, covering regret with a block bound, and completing tasks at a nominal sensing

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.

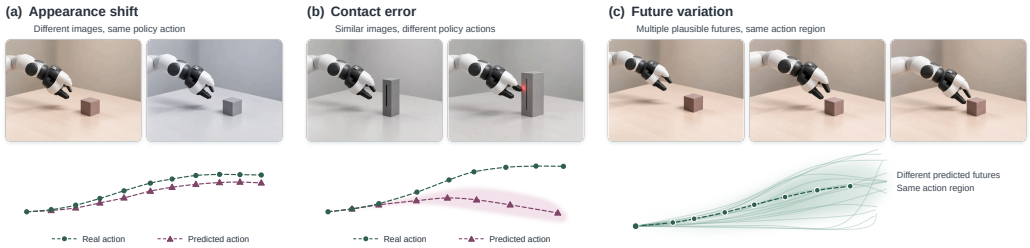

Figure 1: Visual differences and policy responses need not agree. In (a) and (b), each curve compares the action proposed from a real observation (circles) with the action proposed from a predicted observation (triangles). (a) An appearance change alters the image but leaves the action nearly unchanged. (b) A small contact error leaves the images similar but moves the predicted action away from the real one. (c) Several plausible predicted futures (thin lines) remain within one action region.

rate. DREAM-CAL has higher closed-loop success and better harm ranking than the visual head, but loses on four tasks, has weaker ECE, and loses coverage after dynamics shift.

The technical contribution is a policy-conditioned regret target, a learned head for that target, and a horizon rule whose conformal statement accounts for selection within a fixed anchor-horizon family. The statement is marginal over exchangeable trajectory blocks. We evaluate the method across 37 simulation tasks, six model-policy configurations, and four physical tasks. Section 5 reports the results, and Appendix B details the paired-continuation protocol.

2 Related Work

World-Model Reliability. Latent models support planning (Hafner et al., 2019) and policy learning through imagination (Hafner et al., 2020). PETS propagates ensemble uncertainty during planning (Chua et al., 2018); MBPO limits model bias through short training rollouts (Janner et al., 2019). Prediction accuracy alone need not determine control performance (Lambert et al., 2020), motivating value-aware model losses (Farahmand et al., 2017). C3 calibrates dense video uncertainty (Mei et al., 2025), HAUWM structures uncertainty across horizons (Wan et al., 2026), and MiraBench evaluates action following and failure preservation (Yang et al., 2026). We instead estimate the consequence of imagined inputs for a frozen policy, without optimizing that policy or the predictor.

Visual Feedback and Action Execution. Action chunking amortizes policy inference (Zhao et al., 2023; Chi et al., 2023). Sensory value depends on task dynamics and policy proficiency (Krishna et al., 2025). Our paired target measures the cost of replacing feedback with a specified imagination-driven continuation. FFDC verifies predicted dynamics against real observations to choose execution length (Wang et al., 2026); our gate decides before acquiring the next image. KnowNo uses conformal prediction to request human help for ambiguous language-model plans (Ren et al., 2023). Here the requested resource is visual feedback for continuous control, and supervision comes from paired rollout outcomes rather than plan ambiguity.

Calibration for Decisions. Probability calibration measures probability-frequency agreement (Guo et al., 2017); selective prediction studies the risk-coverage trade-off under rejection (Geifman & El-Yaniv, 2019). Conformalized quantile regression calibrates predictive bounds (Romano et al., 2019), whereas conformal risk control and decision theory target losses or downstream decisions (Angelopoulos et al., 2024; Lekeufack et al., 2023). PlanCP calibrates diffusion-model trajectory predictions (Sun et al., 2023). Our calibrated quantity is paired regret, with simultaneous coverage across a fixed horizon family; it does not guarantee conditional safety among accepted actions. Adaptive and weighted methods address changing distributions (Gibbs & Candès, 2021; Barber et al., 2023). Recalibrating our counterfactual target requires new paired continuations, beyond the feedback obtained from one deployed trajectory.

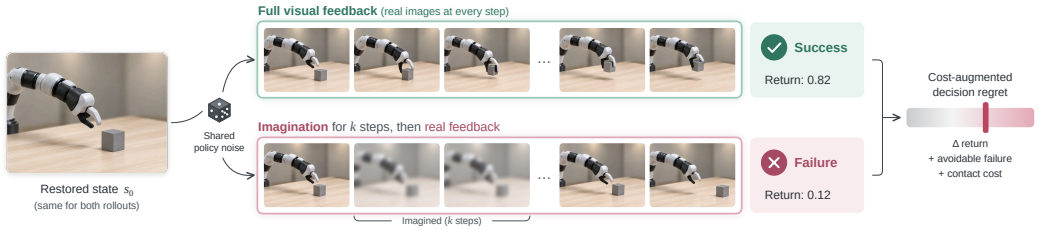

Figure 2: Offline paired supervision. Both rollouts start from the same restored state and run the frozen policy with shared sampling noise. The upper branch receives a real image at every step; the lower branch receives imagined observations for k steps and then returns to real feedback. The return difference, the avoidable-failure indicator and the contact cost of the imagination branch form the cost-augmented regret in Equation 1.

3 Decision Regret

The robot has a real-image history h_t, proprioception p_t and task condition c. A frozen policy π proposes an action chunk $\mathbf{a}_t$ of maximum length $K = 16$. A frozen world model W predicts observations conditioned on this history and the proposed actions. The gate chooses a visual refresh or an interval $k \in \mathcal{K} = \{1, 2, 4, 8, 16\}$ during which new images are withheld. Proprioception, low-level control and policy evaluation on imagined inputs continue during this interval.

A refresh rate is the number of acquired visual updates divided by the number of control steps. It differs from the fraction of candidate segments rejected by an offline score and from the number of policy calls. Our empirical objective is task success at a specified refresh rate. The reference policy receives full visual feedback; it is an evaluated reference, not a mathematical upper bound on performance.

For a state s_t available during offline labelling, branch F follows the reference protocol. Branch $I(k)$ uses imagined observations for k steps and then follows the same reference protocol. The policy parameters and initial state are shared, and random streams are coupled across branches (Figure 2). Let $J_F, J_{I(k)} \in [0, 1]$ be terminal task returns, $A_{t,k}$ indicate failure in $I(k)$ when F succeeds, and $C_{t,k} \in [0, 1]$ be the normalized absolute contact cost of the imagination branch. We define

$$R_{t,k} = J_F - J_{I(k)} + \lambda_f A_{t,k} + \lambda_c C_{t,k}, \qquad \lambda_f = 1, \quad \lambda_c = 0.5. \tag{1}$$

The contact cost is not reference-subtracted: equal task outcomes can still yield positive regret. Negative labels retain improvements after accounting for this cost. The absolute contact term can flag a segment even if refreshing vision would not have prevented its contact; the label is a composite risk, not a claim that every flagged event was avoidable by sensing. Harm is the composite event $Y_{t,k} = \mathbf{1}[R_{t,k} > \epsilon]$, with $\epsilon = 0.2$; because the label includes costs, this threshold is not simply a percentage loss of task return.

The deployed information set contains the acquired history, current proprioception, proposed actions and predictions from W. It excludes the unacquired image, simulator state and paired outcomes. The state and both continuations are required only when constructing labels, as detailed in Appendix B.

4 Policy-Conditioned Calibration

DREAM-CAL converts action-conditioned predictions into evidence about a frozen policy's response, fits a regret head offline, and uses a calibrated upper bound to choose the next visual refresh. Figure 3 shows the online horizon choice together with the offline supervision and calibration on which it relies.

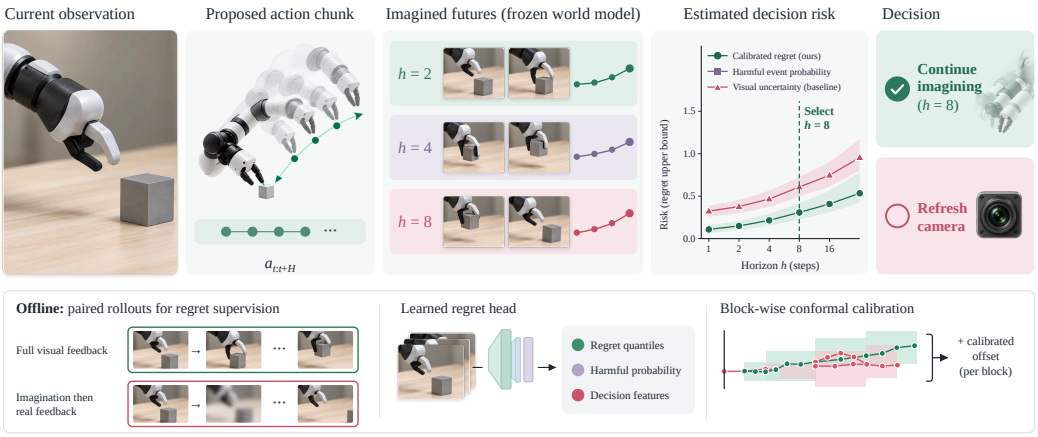

Figure 3: DREAM-CAL overview. Top: the frozen world model predicts futures for the proposed action chunk at candidate horizons, and the learned head turns policy-conditioned features of these futures into a calibrated regret bound for each horizon. The gate continues with imagined observations up to the longest admissible horizon, here eight steps, or refreshes the camera. Bottom: paired rollouts supervise the regret head, and block-wise conformal calibration sets the offset added to its upper quantile.

4.1 Policy-Conditioned Evidence

For each candidate horizon, the world model draws $M = 4$ observation sequences at the decision anchor,

$$\hat{o}^{(m)}_{t+1:t+k} \sim W(\cdot \mid h_t, p_t, c, \mathbf{a}_{t,1:k}), \qquad m = 1, \ldots, M. \tag{2}$$

Sample $m = 1$ is designated as the execution future before scoring; all four samples enter the decision features, but only $m = 1$ supplies imagined images to the paired continuation and online controller. Its visual sequence is generated once from the initial proposed action chunk and is not regenerated if later policy actions differ from that chunk. During the selected interval, the policy continues to act on the fixed imagined images and the latest real proprioception, obtained from the simulator state or the robot's joint sensors. The anchor-time probes below use the currently available p_t; future proprioception is not predicted by W.

The same visual variation can affect different policies differently. We therefore probe π on each predicted sequence and retain the induced action distributions, rather than reducing every future to a visual error score. If $\mu_m = \pi(\cdot \mid \hat{h}^{(m)}, p_t, c)$ is a probe distribution, a summary feature is

$$d_{\pi,k} = \frac{2}{M(M-1)} \sum_{m<n} D_a(\mu_m, \mu_n), \tag{3}$$

where D_a compares action chunks using 64 one-dimensional sliced Wasserstein projections ($M \geq 2$). These projection directions estimate pairwise discrepancies and are not 64 separate head-input coordinates. For $M = 1$, there is no future pair: the between-future action-disagreement features are zero-filled without evaluating Equation 3. Features from the single imagined future, including policy response, action margin and visual uncertainty, remain available. Diffusion-policy probes share sampling noise across imagined inputs. For the deterministic L1-action VLA, each μ_m is a point mass at its predicted chunk.

The head also receives local action-margin features and visual uncertainty weighted by policy input sensitivity. Its discrete-stage feature slot is zero-filled for the evaluated policies, which have no stage output. Context encodes proprioception, candidate actions, observation age and horizon. Candidate conditioning lets the same visual state receive different scores for different actions. Visual features retain model-uncertainty evidence when action probes agree. Feature dimensions and probe settings appear in Appendix A.

4.2 Learning Paired Regret

A four-layer multilayer perceptron predicts conditional regret quantiles $\hat{q}_{0.5}(x)$ and $\hat{q}_{0.9}(x)$, harmful probability $\hat{p}(x)$, and an auxiliary failure category. Neither W nor π receives gradients. With inverse-inclusion weight w_i for a stratified training sample, the head objective is

$$\mathcal{L} = \frac{1}{\sum_i w_i} \sum_i w_i \left[\sum_{\tau \in \{0.5, 0.9\}} \rho_\tau(R_i - \hat{q}_\tau(x_i)) + \mathrm{BCE}(Y_i, \hat{p}(x_i)) + 0.3\, \mathrm{CE}(z_i, \hat{z}_i) \right], \quad (4)$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}[u < 0])$ is the pinball loss and z_i is the category label. The quantitative audit is uniform over held-out blocks; its reported event frequencies use direct counts rather than training-sample weights.

The probability and quantile outputs serve different purposes. The former orders harmful events and supports a probability threshold; the latter predicts the magnitude of regret and supplies residuals for an upper bound. An intercept-only probability adjustment is fitted on the disjoint calibration split for the main probability comparisons. Adding a conformal offset to $\hat{q}_{0.9}$ leaves $\hat{p}$ and its ranking unchanged.

4.3 Calibration over Trajectory Blocks

Anchors from one trajectory share dynamics and scene context. We keep those anchors in the same split and calibrate at the block level. For block b with a fixed set of anchors $\mathcal{A}_b$, define

$$S_b = \max_{i \in \mathcal{A}_b,\, k \in \mathcal{K}} \left[R_{i,k} - \hat{q}_{0.9}(x_{i,k}) \right], \qquad U_{t,k} = \hat{q}_{0.9}(x_{t,k}) + q^\star. \quad (5)$$

With n calibration blocks and target miscoverage $\alpha = 0.1$, $q^\star$ is order statistic $\lceil (n+1)(1-\alpha) \rceil$ of their scores, using $+\infty$ if that rank exceeds n. Taking the maximum across the fixed horizon family includes the selection step in the calibrated event.

Proposition 1 (Block Coverage)**.** *Condition on a fitted head and a fixed block-construction protocol. If the n calibration blocks and a test block are exchangeable, Equation 5 gives $\Pr\{R_{i,k} \le U_{i,k}$ for every $i \in \mathcal{A}_{\text{test}},\, k \in \mathcal{K}\} \ge 1 - \alpha$. Consequently, any selected anchor-horizon pair within that test block is covered on the same simultaneous event.*

The result is the conformal rank argument applied to block maxima (Romano et al., 2019); Appendix D provides the proof and defines its scope. It is marginal over calibration and test blocks, rather than conditional on a task, scene or accepted action. It also does not cover a new state distribution induced by an altered controller unless the exchangeability condition continues to hold.

4.4 Visual Refresh Decisions

The gate selects the longest admissible horizon,

$$k_t^\star = \max\{k \in \mathcal{K} : U_{t,k} \le \epsilon,\ \hat{p}(x_{t,k}) \le \delta,\ k \le b_t\}, \quad (6)$$

where b_t is the remaining episode horizon. If the set is empty, the robot acquires a real image. Otherwise it executes with imagined visual inputs until the selected refresh point. The nominal setting uses $\delta = 0.05$; refresh-frontier experiments sweep the gate operating point and compare the resulting sensing rates. The candidate action comes from the frozen policy, not an additional optimizer of world-model reward.

The probability threshold is an operating choice, not a consequence of Proposition 1. In particular, 90% marginal regret coverage does not imply 5% harmful risk among accepted segments. Choosing a long admissible interval is also a local rule: it changes the next state and the opportunities to acquire feedback, so it need not minimize sensing over an entire episode.

5 Experiments

We evaluate closed-loop success, harmful-event ranking and calibration, regret-bound coverage, component ablations, model-policy transfer, and physical-robot outcomes. Wilson intervals in the appendix summarize uncertainty in the observed binary success counts under a binomial model.

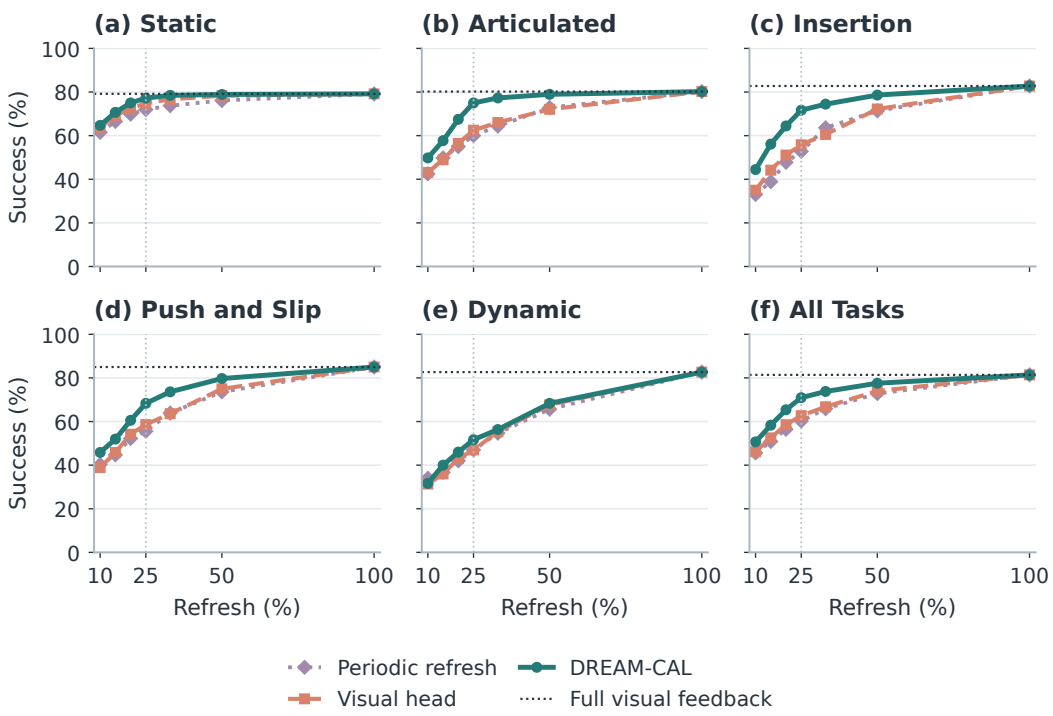

Figure 4: Success across five task families and their pooled total versus realized visual-refresh ratio. DREAM-CAL rates result from sweeping the gate operating point; its nominal 25% point corresponds to the 37-task results reported in the text. Periodic refresh uses fixed schedules; the visual head and full visual feedback provide comparisons.

5.1 Evaluation Protocol

The evaluation contains 37 tasks: 12 static, eight articulated, six insertion, six pushing and slip-recovery, and five dynamic interception tasks. At the nominal 25% operating point, each method has 60 trials per task, or 2,220 trials in total. Binary task completion is the control endpoint. Equal task denominators make the pooled rate identical to the unweighted task mean.

The harm audit has 2,400 prespecified (anchor, k) segments from 480 five-anchor blocks: each block assigns the five horizons one-to-one to its anchors before scores or paired outcomes are observed, giving 480 segments per horizon. Of these, 469 have harmful labels defined by $R_{t,k} > 0.2$. AUROC and average precision assess ranking; equal-width-bin ECE and Brier score assess the probability output. Selective risk counts harmful events among low-score prefixes and does not measure task completion. Block coverage is assessed in a separate 240-block stress design, and selected-decision coverage uses one decision per test block. A contact-sensitive subset contains the six insertion, six push/slip, and five dynamic tasks (1,020 trials per method). Appendix B specifies the paired-continuation and split protocol; Appendix C defines success and harm.

5.2 Closed-Loop Success

At nominal 25% refresh, DREAM-CAL records 1,575/2,220 successes (70.95%) and the visual head 1,392/2,220 (62.70%). The integer-count difference is 183/2,220, or 8.24 percentage points. Insertion and articulated tasks contribute 57 and 60 of those additional successes, respectively; their family gains are 15.83 and 12.50 points. Static and dynamic gains are smaller. Across individual tasks, 31 counts favor DREAM-CAL, two tie, and four favor the visual head (Appendix F). The aggregate therefore does not imply a uniform task-level advantage.

Figure 4 shows task success across visual-refresh rates. The 25% point is used for the detailed 37-task comparison below.

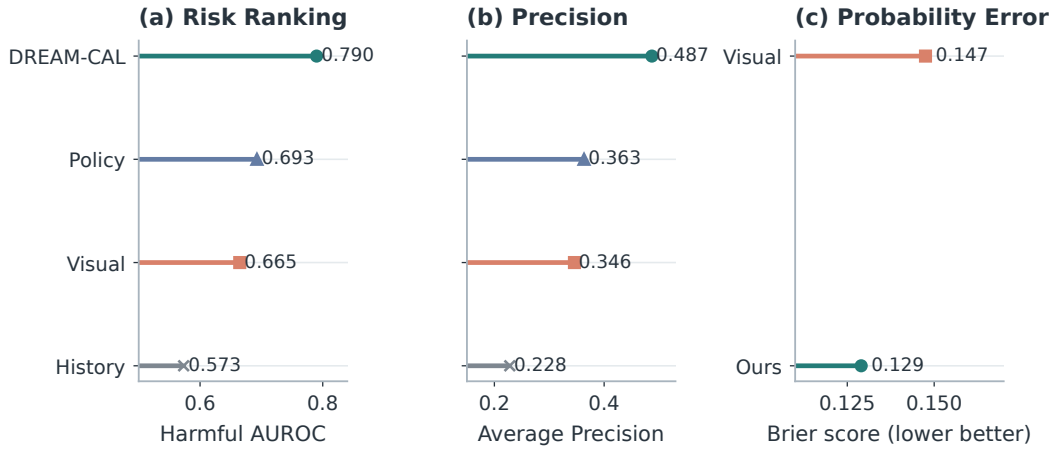

Figure 5: Harmful-event ranking and Brier score on 2,400 evaluated segments. Lower Brier score and higher AUROC or average precision are favorable.

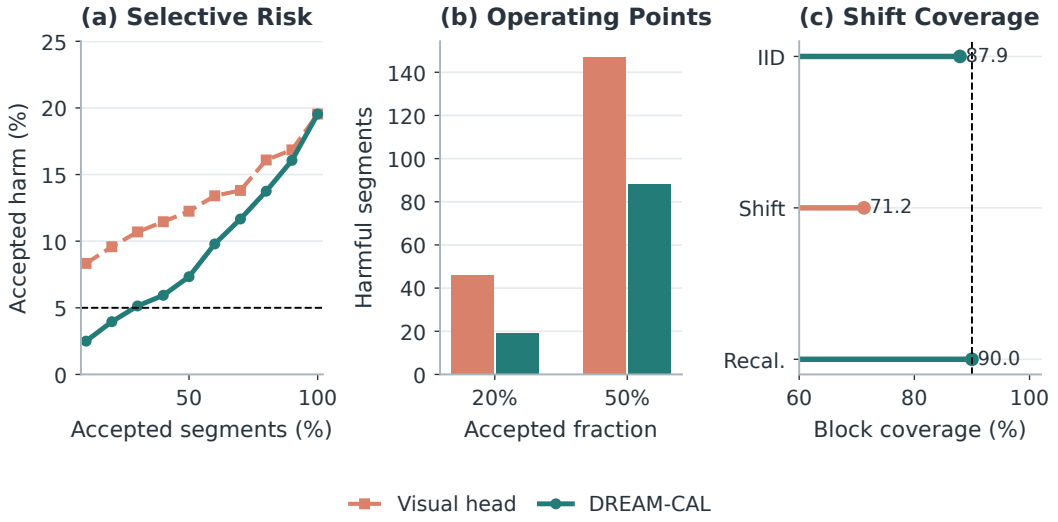

Figure 6: Selective-risk curve, accepted-harm counts at 20% and 50%, and block coverage. The right panel uses 240 blocks per condition and a dashed 90% reference.

5.3 Decision Risk

The segment audit gives DREAM-CAL harmful-event AUROC 0.790 and average precision 0.487, compared with 0.665 and 0.346 for the visual head. The ranking improvement does not uniformly improve probability calibration: Brier score changes from 0.1475 to 0.1290, whereas ECE changes from 0.0089 to 0.0145 in the opposite direction.

At 20% acceptance, the accepted set contains 19/480 harmful segments for DREAM-CAL (3.96%) and 46/480 for the visual head (9.58%). At 50%, the respective counts are 88/1,200 (7.33%) and 147/1,200 (12.25%). At full acceptance, both methods reach the common base rate of 469/2,400 harmful segments. The gate also checks a regret bound, so this score-sorting curve is not its deployed acceptance set.

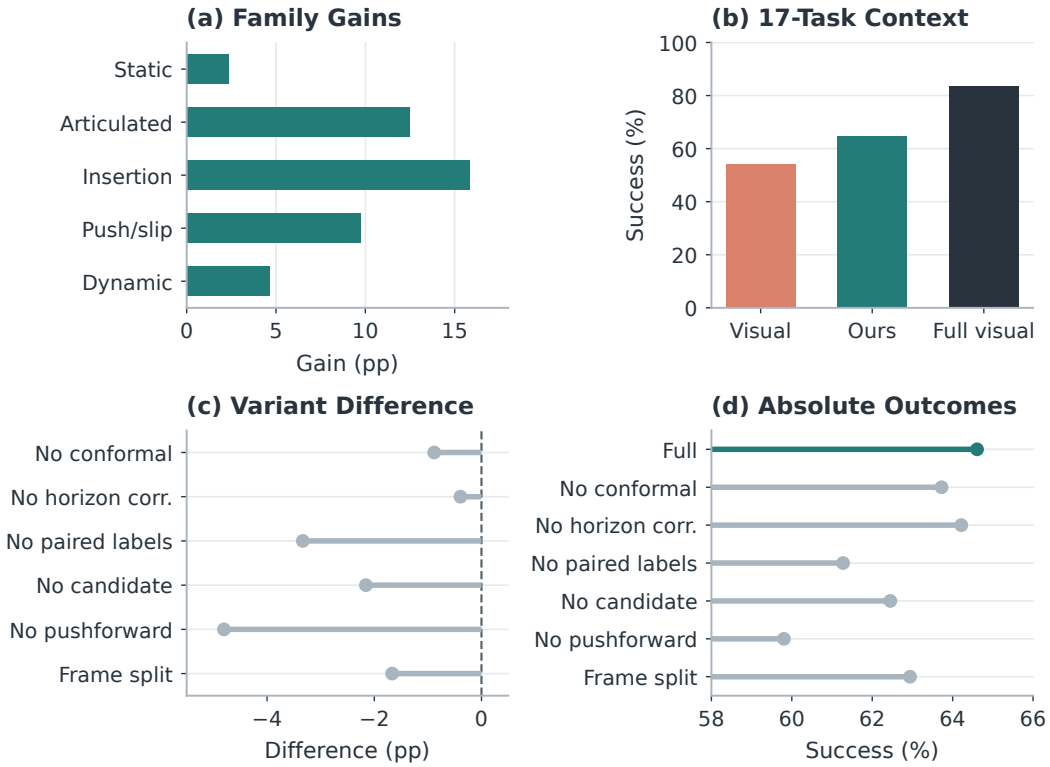

Figure 7: Family gains, 17-task context, and component comparisons. The 17-task panels use 1,020 trials per method.

5.4 Coverage under Shift

The separate block audit finds 211/240 covered IID blocks (87.92%). Holding the original offset under dynamics shift gives 171/240 (71.25%); recalibrated coverage, after 180 new calibration blocks, is 216/240 (90.00%). A distinct selected-decision audit has 222/240 covered decisions (92.50%). These are different statistical units and cannot be pooled. Proposition 1 concerns marginal simultaneous coverage for exchangeable blocks under a fixed construction protocol. These frequencies do not provide a conditional guarantee for accepted decisions after shift. Appendix E separates the block and selected-decision comparisons.

5.5 Components and Transfer

The 17 contact-sensitive tasks account for 659/1,020 DREAM-CAL successes and 553/1,020 visual-head successes; the full-feedback reference has 852/1,020. Within that subset, the no-pushforward, no-paired-label, and no-candidate variants reach 610, 625, and 637 successes. Removing the conformal offset or horizon correction gives 650 and 655. These comparisons support the utility of policy-conditioned features, although changes to the decision rule can also change the realized sensing frequency.

Six 1,020-trial model-policy settings give five positive transfer differences and one loss of one trial relative to their visual heads (Appendix F). The largest gain is 66 trials for Video/DP; Decision-Aware/VLA changes by -1 trial. The transfer Video/DP row uses the default architecture pair but independently instantiated frozen checkpoints, heads and calibration; it is not a reaggregation of the default 17-task trials. The settings should not be pooled with that comparison as repeated runs of one configuration. Appendix G distinguishes changes to features, supervision, and calibration.

5.6 Interpreting the Comparisons

The analyses use different statistical units. Task success is counted over 37 tasks with 60 trials per task; the risk audit uses offline continuation segments; the coverage stress test uses trajectory blocks; transfer settings each use a 17-task model-policy configuration; and the physical study uses four tasks. A success gain cannot be derived directly from a lower segment harm rate or a coverage percentage.

Wilson intervals in the appendix summarize each method's binary success count under a binomial model. Although initial states are matched, the manuscript reports neither the discordant trial counts nor a paired confidence interval or significance test; the 183-success difference is an observed count difference. The risk-ranking curve sorts segment scores, while the deployed gate also checks a regret bound; its accepted set can therefore differ from the score-sorting curve. The conformal result concerns exchangeable blocks and does not extend automatically to shifted dynamics or entire deployed episodes.

Each simulation method is evaluated on the same reset states, with 20 trials under each of three evaluation seeds per task. These seeds index reset trials, not independently trained models. The reported Wilson intervals describe marginal success rates; the paired design would also permit an analysis of which resets switch outcome between controllers, but those discordant counts are unavailable. Accordingly, the task-level wins and the pooled 183-success difference are descriptive comparisons. Likewise, the component variants share a nominal operating setting rather than an exactly matched realized image count. Their success totals measure the resulting complete controllers, including any change in acquisition frequency caused by a modified gate.

5.7 Physical and Computational Scope

The four-task physical study includes 25 trials per task and method. DREAM-CAL totals 66/100 successes, the visual head 61/100, periodic refresh 53/100, and full feedback 76/100 (Appendix I). The five-trial difference is small relative to the 100-trial denominator. The physical models are deployed from simulation without weight updates, and method-level realized visual-refresh ratios are not reported; these outcomes compare complete controllers rather than refresh strategies at a verified equal sensing rate. The measured compute time also limits the interpretation of simulated success: DREAM-CAL's reported 452 ms mean model-compute time exceeds both the nominal 100 ms control-step period and a four-step, 400 ms accounting interval. Appendix H separates model-compute latency from camera acquisition, data transfer, and robot execution.

The robot uses the corresponding simulation task's frozen world model, policy, regret head and calibration parameters. Physical deployment adapts camera calibration, image preprocessing, joint-state normalization and action coordinates deterministically; physical trials do not update model weights. Each task has a measured completion predicate sustained for five control steps, a 60-second timeout and protective stops counted as failures. High-level inference is synchronous while the low-level controller holds its current setpoint. Thus the nominal control period and the compute-only timing do not establish a measured end-to-end decision rate on the robot.

6 Conclusion

DREAM-CAL defines visual refresh through the regret of a frozen policy acting on imagined observations, with a block-calibrated upper bound for a fixed horizon family. Across 37 simulation tasks, its nominal 25% operating point gains 183 successes over a visual-uncertainty head, with the largest family differences in articulated and insertion tasks. The results also show limitations: task losses remain, ranking and probability calibration differ, frozen coverage falls under dynamics shift, and measured decision latency exceeds the target control period, motivating a faster implementation before real-time deployment.

Ethics Statement

Physical experiments use protective stops, which count as failures. Coverage drops under dynamics shift with the original calibration offset, so deployment requires monitoring, recalibration, and protective controls. The reported results should not be interpreted as a general safety guarantee.

Reproducibility Statement

The main text specifies the regret target, learning objective, calibration rule, and evaluation metrics. Appendices A-I document model settings, paired-state restoration, data splits, task-completion criteria, calibration, and computational cost. These details support reproduction of the reported experiments.

AI Use Statement

AI tools assisted with literature search, manuscript organization and writing edits, discussion of experimental design, and code for figures and analysis. The authors are responsible for the experiments, numerical results, references, analyses, and final manuscript.

References

Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal Risk Control. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=33XGfHLtZg.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal Prediction beyond Exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023. doi: 10.1214/23-AOS2276. URL https://doi.org/10.1214/23-AOS2276.

Victor Chernozhukov, Kaspar Wüthrich, and Yinchu Zhu. Exact and Robust Conformal Inference Methods for Predictive Machine Learning with Dependent Data. In *Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 732–749, 2018. URL https://proceedings.mlr.press/v75/chernozhukov18a.html.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin CM Burchfiel, and Shuran Song. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. In *Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.026. URL https://roboticsproceedings.org/rss19/p026.html.

Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep Reinforcement Learning in a Handful of Trials using Probabilistic Dynamics Models. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL https://papers.neurips.cc/paper_files/paper/2018/hash/3de568f8597b94bda53149c7d7f5958c-Abstract.html.

Amir-Massoud Farahmand, Andre Barreto, and Daniel Nikovski. Value-Aware Loss Function for Model-Based Reinforcement Learning. In *International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1486–1494, 2017. URL https://proceedings.mlr.press/v54/farahmand17a.html.

Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1050–1059, 2016. URL https://proceedings.mlr.press/v48/gal16.html.

Yonatan Geifman and Ran El-Yaniv. SelectiveNet: A Deep Neural Network with an Integrated Reject Option. In *International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2151–2159, 2019. URL https://proceedings.mlr.press/v97/geifman19a.html.

Isaac Gibbs and Emmanuel Candès. Adaptive Conformal Inference under Distribution Shift. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2106.00170.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. In *International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330, 2017. URL https://proceedings.mlr.press/v70/guo17a.html.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning Latent Dynamics for Planning from Pixels. In *International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2555–2565, 2019. URL https://proceedings.mlr.press/v97/hafner19a.html.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to Control: Learning Behaviors by Latent Imagination. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1lOTC4tDS.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. URL https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Michael Janner, Justin Fu, Marvin Zhang, and Sergey Levine. When to Trust Your Model: Model-Based Policy Optimization. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/5faf461eff3099671ad63c6f3f094f7f-Abstract.html.

Moo Jin Kim, Chelsea Finn, and Percy Liang. Fine-Tuning Vision-Language-Action Models: Optimizing Speed and Success. In *Robotics: Science and Systems*, 2025a. doi: 10.15607/RSS.2025.XXI.017. URL https://www.roboticsproceedings.org/rss21/p017.html.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan P Foster, Pannag R 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 of *Proceedings of Machine Learning Research*, pp. 2679–2713, 2025b. URL https://proceedings.mlr.press/v270/kim25c.html.

Arjun Krishna, Edward S. Hu, and Dinesh Jayaraman. The Value of Sensory Information to a Robot. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e1126028d9f1f69c13571ec462084d31-Abstract-Conference.html.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.neurips.cc/paper_files/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html.

Nathan Lambert, Brandon Amos, Omry Yadan, and Roberto Calandra. Objective Mismatch in Model-Based Reinforcement Learning. In *Learning for Dynamics and Control*, volume 120 of *Proceedings of Machine Learning Research*, pp. 761–770, 2020. URL https://proceedings.mlr.press/v120/lambert20a.html.

Jordan Lekeufack, Anastasios N. Angelopoulos, Andrea Bajcsy, Michael I. Jordan, and Jitendra Malik. Conformal Decision Theory: Safe Autonomous Decisions from Imperfect Predictions. *arXiv preprint arXiv:2310.05921*, 2023. URL https://arxiv.org/abs/2310.05921.

Zhiting Mei, Tenny Yin, Micah Baker, Ola Shorinwa, and Anirudha Majumdar. World Models That Know When They Don't Know – Controllable Video Generation with Calibrated Uncertainty. *arXiv preprint arXiv:2512.05927*, 2025. URL https://arxiv.org/abs/2512.05927.

Allen Z. Ren, Anushri Dixit, Alexandra Bodrova, Sumeet Singh, Stephen Tu, Noah Brown, Peng Xu, Leila Takayama, Fei Xia, Jake Varley, Zhenjia Xu, Dorsa Sadigh, Andy Zeng, and Anirudha Majumdar. Robots That Ask For Help: Uncertainty Alignment for Large Language Model Planners. In *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pp. 661–682, 2023. URL https://proceedings.mlr.press/v229/ren23a.html.

Yaniv Romano, Evan Patterson, and Emmanuel Candès. Conformalized Quantile Regression. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/5103c3584b063c431bd1268e9b5e76fb-Abstract.html.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=St1giarCHLP.

Jiankai Sun, Yiqi Jiang, Jianing Qiu, Parth Nobel, Mykel J. Kochenderfer, and Mac Schwager. Conformal Prediction for Uncertainty-Aware Planning with Diffusion Dynamics Model. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/fe318a2b6c699808019a456b706cd845-Abstract-Conference.html.

Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A Physics Engine for Model-Based Control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033, 2012. doi: 10.1109/IROS.2012.6386109. URL https://ieeexplore.ieee.org/document/6386109/.

Shenghua Wan, Le Gan, and De-Chuan Zhan. Learning to Be Uncertain: Pre-training World Models with Horizon-Calibrated Uncertainty. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ad222e6e38e26e3f00d85829b9e02f43-Abstract-Conference.html.

Rui Wang, Yue Zhang, Jiehong Lin, Kuncheng Luo, Jianan Wang, Zhongrui Wang, and Xiaojuan Qi. When to Trust Imagination: Adaptive Action Execution for World Action Models. *arXiv preprint arXiv:2605.06222*, 2026. URL https://arxiv.org/abs/2605.06222.

Tianzhuo Yang, Zihan Shen, Zirui Mi, Zhaoyi Zhang, Jiayi Zhou, Jiaming Ji, Juntao Dai, Jiawei Chen, Boyuan Chen, and Yaodong Yang. MiraBench: Evaluating Action-Conditioned Reliability in Robotic World Models. *arXiv preprint arXiv:2605.29360*, 2026. URL https://arxiv.org/abs/2605.29360.

Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware. In *Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.016. URL https://roboticsproceedings.org/rss19/p016.html.

Yuke Zhu, Josiah Wong, Ajay Mandlekar, and Roberto Martín-Martín. robosuite: A Modular Simulation Framework and Benchmark for Robot Learning. *arXiv preprint arXiv:2009.12293*, 2020. URL https://arxiv.org/abs/2009.12293v1.