NEURAL INFORMATION PROCESSING SYSTEMS

ROBOT LEARNING WITH WORLD MODELS: CAPABILITIES, FRONTIERS, AND CHALLENGES

HALT: Certifying Trust Horizons in Action-Conditioned World Models

How long is a world model's imagination actually valid? Ask it, per state.

Noor Islam S. Mohammad[1*] Md. Faisal Sheikh[2†] Jakaria Habib[3†]

[1]Istanbul Technical University [2]North South University [3]Pabna University of Science and Technology

[*]Corresponding author · [†]Equal contribution · islam23@itu.edu.tr

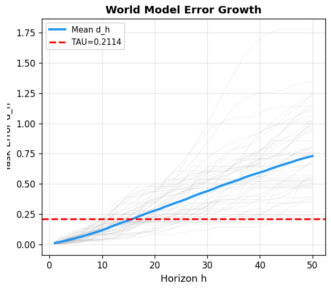
The imagination horizon is a guess that never gets checked

Every latent-imagination agent rolls its model forward H steps and plans inside that rollout. H is chosen **once**, by hand, and applied to every state.

But how far a model stays valid is **state-, action- and context-dependent**. A rollout can look perfectly realistic and still have left the real dynamics ten steps ago.

THE FAILURE WE NAME AND MEASURE

$$\mathrm{SFR}(\tau, \kappa) = \mathbb{P}\big[\, \pi_{\mathrm{app}} \geq \kappa, \;\; \mathrm{TTCD}(\tau) \leq H_{\mathrm{task}} \,\big]$$



Grey: individual rollouts; red: the calibrated tolerance. Where a rollout crosses is state-dependent.

The claim. Judging a world model on whether its futures *look* real misses what a robot needs to know: are they *true*, and for how long?

What HALT needs, stated before what it gives

The guarantee is real only when the interface is available — so we put the interface first

Requirement	Needed for	If unavailable
Stochastic / ensemble rollouts	disagreement feature	coverage ~ 0.91
Differentiable task readout	Jacobian features	coverage ~ 0.91
Task state at calibration	fitting $\hat{q}_h$, $\hat{\lambda}$	no certificate

Task state at *deployment* is needed only for the adaptive variant; without it, static conformal still applies.

Error is measured in **task units, never pixels**:

$$d_h = \|y_{t+h} - \hat{y}_{t+h}\|_\Sigma\,, \qquad \Sigma = \mathrm{diag}(\sigma_1^{-2}, \ldots, \sigma_{d_y}^{-2})$$

σ_i is the tolerance of coordinate i. So $\tau = 1$ means *at the edge of what the task allows* — one threshold, every task.

What this paper does *not* claim.

HALT is not a drop-in wrapper for arbitrary video models, gives no finite-sample guarantee under adversarial shift, and does not remove the need for a task-state estimator when ACI is used.

Video-only models satisfy the first requirement but not the second: we run them with the Jacobian family removed and **report the coverage drop**.

Monotone by construction, calibrated by one scalar

1 — MONOTONE QUANTILE HEAD

$$\hat{q}_h(\varphi_t) = \sum_{j=1}^{h} \mathrm{softplus}\big(\delta_j(\varphi_t)\big)$$

Non-decreasing in h **architecturally**, not by penalty. So the certified horizon is a **single threshold crossing** — no post-hoc isotonic repair.

2 — ONE SCALE-FREE MAX-SCORE

$$R^{(i)} = \max_{1 \le h \le H_{\max}} \frac{d_h^{(i)}}{\hat{q}_h\big(\varphi^{(i)}\big)},$$

$$\hat{\lambda} = \mathrm{Quantile}_{\lceil (n+1)(1-\alpha) \rceil / n}\big(\{R^{(i)}\}_{i=1}^{n}\big).$$

The **max over h buys simultaneity for free** — one scalar, calibrated once, where a Bonferroni correction across horizons would be loose and non-simultaneous.

$$H^\star(\varphi_t; \tau, \alpha) = \max\big\{h \le H_{\max} \ : \ \hat{\lambda}\, \hat{q}_h(\varphi_t) \le \tau\big\}$$

No retraining of the world model.
No per-environment threshold tuning.

If $H^\star < H_{\min}$, the controller abstains.
Slow, retract, re-observe, or ask for help.

The certificate, and what it buys the controller

VALIDITY (PROP. 4.1) — SPLIT CONFORMAL APPLIED TO THE SCALAR R

$$\mathbb{P}\big[\ \forall h \le H^\star(\varphi_t):\ \|y_{t+h} - \hat{y}_{t+h}\|_\Sigma \le \tau\ \big] \ \ge\ 1-\alpha$$

Distribution-free, finite-sample, and **simultaneous over the whole horizon** — the max over h buys simultaneity without the union-bound loss.

WHY A CONTROLLER SHOULD CARE (PROP. 5.1)

$$\big|V^\pi_{\mathcal{M}} - \hat{V}^\pi_{\mathcal{M}_\theta}\big| \le \underbrace{L_r \tau \tfrac{1-\gamma^{H^\star}}{1-\gamma}}_{\text{certified local tolerance}} + \gamma^{H^\star}\varepsilon_V + \underbrace{\tfrac{2\alpha R_{\max}}{1-\gamma}}_{\text{price of the certificate}}$$

Read the three terms.

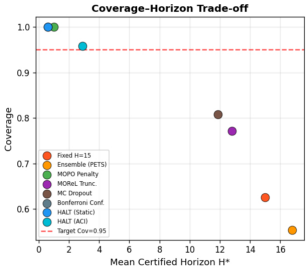
The classical simulation lemma pays global model error compounding. Term one pays only the **certified local tolerance** τ.

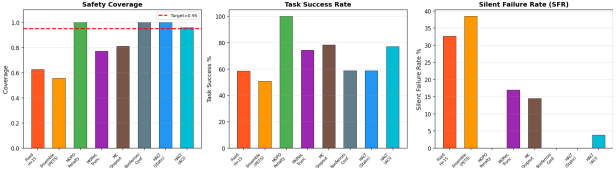
Terms one and two trade off, so an **optimal horizon exists** — state-dependent, untrackable by any hyperparameter, and HALT picks it per control step.

Term three is the honest price of a probabilistic certificate.

Coverage held, horizon shortened, task success up

Five MuJoCo tasks × ten seeds × three world-model classes (DreamerV3, TD-MPC2, 5-model ensemble)





Left: fixed horizons over-trust; pessimistic baselines buy coverage by collapsing the horizon. Right: coverage against the 0.95 target, task success, and silent failure rate.

within 1%
of the 0.95 target, every setting

$\bar{H}^\star \approx 5$ **vs 15**
usable horizon, contact-rich

$+15$ **to** 30%
task success over fixed-horizon

-40 **to** 60%
unsafe contacts

Two places the textbook conformal argument is strained

We expose both rather than overclaim — this is the part a referee should read first

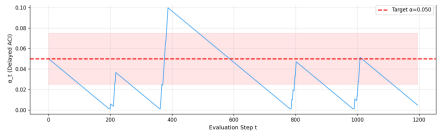
1. The planner breaks exchangeability. Closed-loop MPC *optimises* against the model-predicted objective, so it preferentially selects actions along which the model is most wrong. The executed rollout is no longer drawn from the calibration distribution.

R1 — planner-conditional calibration. Calibrate on rollouts the planner itself produced. Coverage is preserved; $\hat{\lambda}$ rises 1.4–$1.8\times$ and $H^\star$ shortens by 28–41%.

That shortening is **the honest cost of closed-loop planning**, and it is the number an engineer needs.

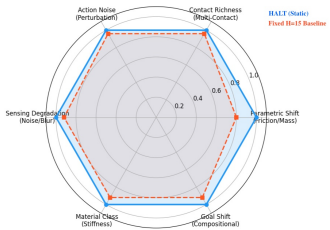
2. The feedback arrives late. ACI needs a miss indicator, but a rollout launched at s is only scored at $s + H_s^\star$: the feedback stream has a variable lag $L_t \le L_{\max}$.

$$\left| \frac{1}{T} \sum_t \mathrm{miss}_t - \alpha \right| \le \frac{\alpha_1 + \eta(1 + L_{\max})}{\eta T} \xrightarrow[T \to \infty]{} 0$$



α_t adapting online under shift; the band is ± 0.03 around target. $L_{\max}$ is the entire price of delay — and the bound is asymptotic, not finite-sample.

Stress-tested on six axes — and what it still cannot do



PHANTASM: six orthogonal stress axes, graded severity. Blue, HALT; orange, fixed $H = 15$, degrading badly under parametric and sensing shift.

LIMITS, STATED PLAINLY

- **Not universal.** Video-only models lose the Jacobian family: coverage drops $0.96 \to 0.92$, and the horizon with it.
- R1 restores exchangeability only when validation resembles deployment; under severe sim-to-real shift neither remedy preserves a guarantee.
- ACI needs y_{t+h} at deployment — rarely available outside simulation or mocap — and its bound is $O(1/T)$ asymptotic.
- All results are in simulation. No real-robot transfer claim.

Cost: 3.2 ms per control step; 12.3 ms total against 9.1 ms for fixed-horizon MPC — inside a 50–100 Hz loop.

NEURAL INFORMATION PROCESSING SYSTEMS

Three things to take away

1 World models are judged on whether their futures *look* real. A robot needs to know whether they are **true, and for how long**. That is a different question, and it is measurable.

2 HALT turns the imagination horizon from a **hyperparameter the engineer guesses** into a **quantity the system reports** — certified, state-dependent, distribution-free, in task-relevant units, with no retraining.

3 The guarantee holds for models that expose the required interface, and we are explicit about where it strains: **planner-induced shift** and **deployment-time ground truth**.

PHANTASM and the six action-conditioned metrics are released as tools for this question:

Code and project page linked from the paper

islam23@itu.edu.tr