NEURONSIFTER: AN AGENT-IN-TWIN FOR PRIORITIZING CENTRAL NERVOUS SYSTEM INTERVENTIONS WITH WORLD ACTION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Effective central nervous system (CNS) intervention requires more than predicting biological responses: a scientific system must determine which intervention, observation, or additional evidence can improve a decision under uncertainty. However, in partially observed neuronal microenvironments, identical predicted outcomes may correspond to different exposure states, biological mechanisms, and intervention risks. We propose NEURONSIFTER, a decision-centric *Agent-in-Twin* framework, where an AI agent operates with an evolving biological digital twin rather than a fixed predictive model. NEURONSIFTER compiles interventions into state-conditional target-occupancy representations, propagates them through a mechanism-grounded *World Action Model (WAM)*, and selects informative measurements based on their expected reduction in decision uncertainty. A shared probabilistic belief state couples intervention simulation, observation selection, and outcome assimilation, enabling the digital twin to continuously refine its decision-relevant representations through externally verified evidence. We evaluate NEURONSIFTER in Alzheimer's disease intervention scenarios using mechanistic neuronal microenvironment simulations and retrospective clinical endpoint references. The evaluation framework examines adaptive evidence acquisition, uncertainty-aware intervention planning, and robustness under model discrepancy through controlled environments and independent comparisons. NEURONSIFTER demonstrates how an AI system can actively identify missing evidence, refine biological hypotheses, and optimize intervention decisions under limited experimental budgets. Our work establishes a decision-oriented framework for adaptive biological digital twins, enabling AI systems to move beyond predicting intervention outcomes toward actively determining which evidence is required to guide therapeutic decisions. Code is available at `https\protect\protect\leavevmode@ifvmode\kern+.2222em\relax//anonymous.4open.science/r/NeuronSifter`.

1 INTRODUCTION

Designing central nervous system (CNS) interventions requires identifying which biological mechanism to target, which intervention to pursue, and which evidence can improve the decision. In neuronal microenvironments, observed phenotypes emerge from coupled processes involving transport, molecular interactions, inflammation, clearance, and cellular adaptation. Distinct biological mechanisms may generate similar trajectories while leading to different intervention strategies. A scientific system should therefore reason over competing mechanistic explanations, simulate candidate interventions, and actively acquire evidence that improves intervention decisions.

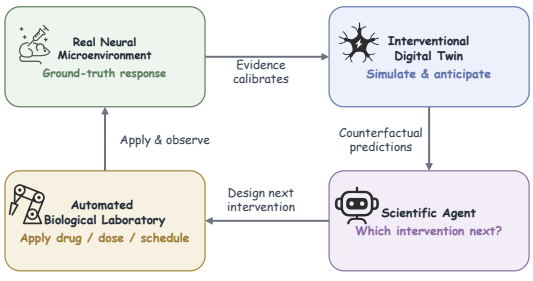


Figure 1: **Twin-in-the-loop intervention optimization for CNS.** An Agent-in-Twin system couples a real neural microenvironment with an interventional digital twin.

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.

Biological digital twins provide a computational foundation for this process by integrating mechanistic simulation with counterfactual reasoning. However, effective intervention design requires more than accurate prediction: the twin must maintain a coherent representation of biological states, connect uncertainty to experimental priorities, and continuously incorporate new evidence. As illustrated in Figure 1, a twin-in-the-loop system couples biological modeling with adaptive evidence acquisition, allowing computational hypotheses and intervention strategies to evolve together.

We introduce NEURONSIFTER, a decision-centric *Agent-in-Twin* framework for CNS intervention selection. A scientific agent operates with a revisable biological twin to compare intervention programs, acquire informative evidence under finite budgets, and optimize decisions through iterative refinement. A unified biological posterior connects these processes: occupancy-based compilation maps interventions to biological inputs, the World Action Model predicts multi-scale trajectories, and decision-focused acquisition selects observations according to their expected impact on intervention outcomes. Typed observations update the posterior, enabling the agent and twin to jointly improve subsequent intervention strategies.

Our contributions are:

- **A decision-centric formulation.** We formulate CNS intervention discovery as a belief-state decision problem that jointly integrates intervention simulation, biological uncertainty, evidence acquisition, and adaptive planning.
- **A shared biological architecture.** We develop an Agent-in-Twin framework that couples occupancy-based intervention representations, World Action Models, posterior updating, and evidence-driven refinement within a unified computational loop.
- **A decision-oriented evaluation framework.** We establish evaluations that compare predictive modeling, intervention selection, and evidence acquisition through transfer scenarios, budget-aware planning, and uncertainty-aware decision analysis.

2 RELATED WORK

Biological twins and uncertainty. Mechanistic AD models couple aggregation, clearance, transport, and neuronal dynamics (Przekwas et al., 2026; Stefanovski et al., 2019). Learned twins predict spatial biomarkers (Wang et al., 2026) and sparse longitudinal transitions (Huang et al., 2026). PIM-BrainTwin models compensatory capacity (Nonn et al., 2026); identifiability and coupling remain broader challenges (Montagud, 2026). Prognostic twins support trial adjustment (Andrews et al., 2026). Our joint exposure–mechanism posterior guides prospective measurement choice. Spatial controls and clinical evaluation appear in Table 31 and Appendix A.4.3.

Adaptive design and planning. EIG/BALD, DAD/iDAD/Step-DAD, and decision-aware BED optimize information or decision targets (Houlsby et al., 2011; Foster et al., 2021; Ivanova et al., 2021; Hedman et al., 2025; Huang et al., 2024a). BAX/PS-BAX target computable properties (Neiswanger et al., 2021; Cheng et al., 2024a); MF-MES and cost-aware BO account for fidelity and cost (Takeno et al., 2020; Xie et al., 2024); BAMCP searches Bayes-adaptive continuations (Guez et al., 2012). We connect these principles to biological posterior updates and selective acceptance; matched planners isolate finite-budget numerical approximation.

3 PROBLEM FORMULATION

We study intervention selection when the same observed neuronal phenotype is compatible with different exposure states and biological mechanisms. These explanations can predict similar untreated trajectories yet favor different interventions. The agent must therefore decide both *what to measure* and *when to commit*, under a finite evidence-acquisition budget. Sparse longitudinal AD twins already support uncertainty-aware what-if trajectories (Huang et al., 2026); here, measurement choice and terminal intervention loss are explicit parts of the belief-state decision problem.

Scientific state and belief. Let $\boldsymbol{\xi}_t = (\mathbf{u}_t, m, \boldsymbol{\theta}, \boldsymbol{\rho}_t)$ contain the microenvironment state, mechanistic regime, biological parameters, and tool reliability. The graph $\mathbf{G}$ specifies biological interactions and

transport. Evidence $\mathcal{E}_t$ induces a belief b_t over this joint state; dependence between exposure and response is retained because it can change intervention rankings. We use t for decision rounds and $h = 0, \ldots, H$ for time within a predicted biological trajectory.

Interventions, measurements, and decisions. A world action $a \in \mathcal{A}^{\mathrm{w}}$ specifies a target program, mechanism, dose, route, schedule, duration, and context. It changes the system through $\mathsf{P}_{\Phi(a,\boldsymbol{\xi}_t,\mathbf{G})}$, where Φ compiles the program into state-conditional biological inputs. An information action $q \in \mathcal{A}^{\mathrm{i}}$ selects an observation channel O_q and returns a typed outcome containing measurements, uncertainty, completion status, cost, and provenance. Pure information actions leave biology unchanged; invasive assays include an explicit world-action component. A terminal decision $d \in \mathcal{D}_{\mathrm{dec}}$ recommends a supported intervention or defers under a predefined loss.

Decision risk and the sequential objective. For intervention a, $\ell(a, \boldsymbol{\xi})$ is its expected trajectory loss conditional on the latent state, including its intervention cost. The current Bayes risk is

$$\mathcal{R}(b) = \min_{d \in \mathcal{D}_{\mathrm{dec}}} \mathbb{E}_{\boldsymbol{\xi} \sim b}[\ell(d, \boldsymbol{\xi})]. \tag{1}$$

An admissible policy $\pi^{\mathrm{AIT}}(\cdot \mid b_t, B_t)$ selects actions or stops using only observed history and remaining budget B_t. If T is its stopping round, the objective is

$$J(\pi) = \mathbb{E}_{\pi}\left[\ell(d_T, \boldsymbol{\xi}_T) + \sum_{t<T} \left(\lambda_c \mathsf{c}_t + \lambda_f \mathbf{1}[\zeta_t \neq \text{success}]\right)\right]. \tag{2}$$

Here c_t is the realized acquisition cost, including partial and failed calls. Loss weights, support rules, and deferral costs are fixed before evaluation. A predictive model supplies possible outcomes; solving Equation (2) additionally requires an observation model and a policy that values evidence by its effect on the eventual decision.

4 METHOD

NEURONSIFTER implements $\pi^{\mathrm{AIT}}(\cdot \mid b_t, B_t)$ from Equation (2) by sharing a biological posterior across intervention prediction, evidence acquisition, and belief revision (Figure 2). The hypotheses explaining the observations therefore also determine predicted responses and measurement value.

4.1 JOINT BIOLOGICAL BELIEF

Uncertain exposure and uncertain biological responsiveness can explain the same observed phenotype while implying different treatment choices. We therefore represent the state $\boldsymbol{\xi}_t = (\mathbf{u}_t, m, \boldsymbol{\theta}, \boldsymbol{\rho}_t)$ using weighted joint particles,

$$b_t = q_\phi(d\boldsymbol{\xi} \mid \mathcal{E}_t, \mathbf{G}) = \sum_{j=1}^{J} w_{tj}\, \delta_{\boldsymbol{\xi}_t^{(j)}}(d\boldsymbol{\xi}). \tag{3}$$

An amortized encoder scores particles from observed evidence, and normalized scores determine their weights. Updating and resampling whole tuples preserves exposure–response dependence.

Observation likelihoods match the available modalities: negative-binomial models describe RNA counts, log-normal models describe positive concentrations, and heteroscedastic Gaussians describe morphology and electrophysiology. Source effects enter the observation model, while missing fields are marginalized (Appendix C.2). Candidate interventions use the same particles and paired random draws, allowing their predicted losses to be compared under common biological uncertainty.

4.2 FROM INTERVENTION PROGRAMS TO PREDICTIVE TRAJECTORIES

Biological responses depend on the exposure and target engagement produced by an intervention, which vary with the latent state. For each particle, the compiler produces time-dependent exposure, occupancy, stimulation, environmental inputs, and support masks. The WAM predicts the resulting

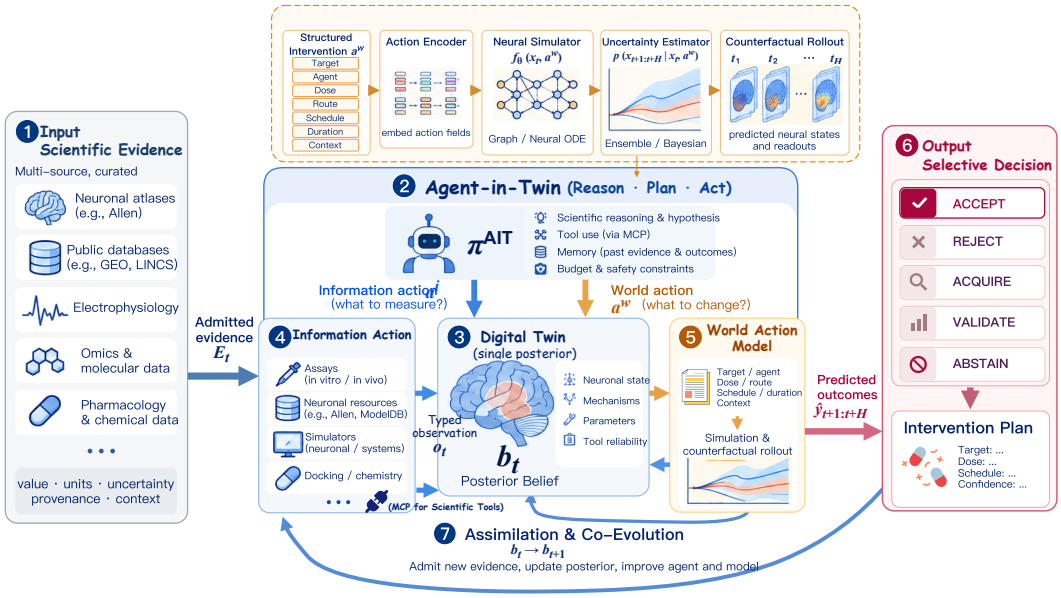


Figure 2: **NEURONSIFTER: decision-centric Agent-in-Twin intervention selection.** A joint biological posterior conditions occupancy compilation, trajectory prediction, and measurement valuation. Typed outcomes update that posterior before replanning; a calibrated terminal gate determines whether the selected intervention has sufficient support.

trajectory conditional on these fields:

$$\boldsymbol{\omega}^a_{0:H} = \Phi(a, \boldsymbol{\xi}^{(j)}_t, \mathbf{G}), \quad \omega^{\text{target}}_{g,h} = \frac{C^{\text{free}}_{g,h}}{C^{\text{free}}_{g,h} + K_{d,g}}, \tag{4}$$

$$X^a_{1:H} \sim p_\theta(\cdot \mid X_0, \boldsymbol{\omega}^a_{0:H}, \boldsymbol{\xi}^{(j)}_t, \mathbf{G}). \tag{5}$$

For target g, $C^{\text{free}}_{g,h}$ denotes free concentration and $K_{d,g}$ the dissociation constant. The pharmacokinetic (PK) model maps dose, route, schedule, and particle-specific transport to exposure. Equilibrium binding or the corresponding kinetic model converts exposure to occupancy; support masks identify unavailable mappings.

To model dynamics across time scales, we quantize multiscale coefficients of $X^a - X_0$ and learn their distribution by masked diffusion (Sahoo et al., 2024). The denoiser conditions on occupancy, posterior context, graph spectral features, and physical and diffusion time. Attention exchanges information across tokens, while ordered residual experts update the representation through

$$h^{(e)} = h^{(e-1)} + \alpha_e \pi_e F_e(h^{(e-1)}, C), \qquad e = 1, \ldots, E. \tag{6}$$

Starting from incoming tokens $h^{(0)}$, expert F_e uses context C, routing weight π_e, and residual scale α_e. Passing updated features between experts lets later transformations respond to earlier changes. Reverse diffusion progressively reveals coefficient tokens; inversion adds X_0 to preserve the initial condition. Sampling across particles and diffusion draws propagates both biological uncertainty and conditional trajectory variation (Appendix C.4).

To connect prediction to the loss in Equation (1), the AD setting uses

$$\mathcal{L}_{\text{CNS}}(a, X^a) = \sum_{h,r} \Delta_h v_r [\lambda_A \bar{A}^a_{hr} + \lambda_\tau \bar{P}^a_{hr} + \lambda_I \bar{I}^a_{hr} + \lambda_E \bar{E}^a_{hr}] + \lambda_D \mathcal{C}_{\text{int}}(a). \tag{7}$$

The normalized channels measure amyloid, tau pathology, inflammation, and circuit dysfunction; Δ_h and v_r weight time and region, and $\mathcal{C}_{\text{int}}$ measures intervention cost. Averaging over rollouts and particle weights estimates $\mathbb{E}_{b_t}[\ell(a, \boldsymbol{\xi})]$. Normalization and preference weights are fixed before evaluation.

4.3 DECISION-FOCUSED EVIDENCE ACQUISITION

The agent values a measurement through the decisions it enables. For an information-only query q, its predictive outcome law averages the observation channel over b. Each hypothetical outcome passes through the same admission and update rules used at execution. The expected reduction in Bayes risk and its net utility are

$$\mathrm{VOI}(q; b) = \mathcal{R}(b) - \mathbb{E}_{\mathbf{o}\sim p(\cdot|b,q)}[\mathcal{R}(\mathsf{T}_q(b, \mathbf{o}))], \tag{8}$$

$$\mathcal{U}(b, q) = \mathrm{VOI}(q; b) - \lambda_c \mathbb{E}[\mathsf{c} \mid b, q] - \lambda_f \mathbb{P}(\zeta \neq \text{success} \mid b, q). \tag{9}$$

Recomputing the minimum risk after assimilation makes acquisition sensitive to intervention ambiguity: uncertainty matters through its effect on the best admissible decision. Partial and failed outcomes contribute through their retained evidence, realized costs, and failure penalties.

Bounded lookahead adds expected continuation value to Equation (9), allowing complementary measurements to be useful jointly (Appendix C.6). The agent executes the best feasible continuation when its value exceeds stopping, then replans with the updated belief and remaining budget. Queries with a world-action component additionally propagate the biological transition before assimilating their outcomes (Algorithm 1).

The shared posterior also links query valuation to intervention comparison. With a common observation channel, decision set, and bounded loss $0 \leq \ell \leq L$, Theorem 4 gives $|\,\mathrm{VOI}(q; b) - \mathrm{VOI}(q; b')| \leq 2L\,\mathrm{TV}(b, b')$. Using the same belief keeps these computations tied to the same biological hypotheses.

4.4 EVIDENCE ASSIMILATION AND SELECTIVE ACCEPTANCE

New evidence updates the biological hypothesis that generated both the intervention forecasts and the query scores. After any world-action transition, the retained outcome $\widetilde{\mathbf{o}}_{t+1}$ updates particle weights through

$$b_{t+1} = \mathsf{T}(b_t, a_t, q_t, \mathbf{o}_{t+1}), \qquad w_{t+1,j} \propto w^-_{tj}\, \overline{p}_{q_t}(\widetilde{\mathbf{o}}_{t+1} \mid \boldsymbol{\xi}^{(j)}_{t+1}). \tag{10}$$

Here w^-_{tj} are predicted weights. The retained likelihood accounts for admitted measurements, censoring, and modeled status events. Failed calls consume budget and update tool reliability; biological updates require admitted values or a validated event likelihood (Appendix C.7). Every candidate is then reevaluated under the revised belief.

Stopping ends acquisition and triggers a separate acceptance decision. The Twin Reliability Index (TRI) aggregates calibration, evidence coverage, PK/PD consistency, ranking stability, and distributional support. The selected intervention a^* is accepted only when

$$\mathrm{accept}(a^*, b) = \mathbf{1}\Big[\widehat{U}_\delta(a^*, b) \leq \tau, \quad \mathrm{TRI}(b, a^*) \geq \tau_r, \quad S(a^*, b) = 1\Big]. \tag{11}$$

S enforces feasibility and support. The upper-loss bound $\widehat{U}_\delta$, with nominal coverage $1 - \delta$, is calibrated on independent episodes after freezing the policy and stopping rule. Thresholds τ and τ_r specify acceptable loss and minimum reliability; failure of any gate condition yields deferral (Appendix C.8).

Learning and execution. Posterior parameters are learned through modality likelihoods and variational regularization. WAM losses combine masked reconstruction, CRPS, physical constraints, and paired action contrasts. Acquisition heads regress Monte Carlo utility and continuation targets on out-of-fold posteriors; update consistency aligns amortized inference with retained-evidence assimilation. Disjoint calibration groups set terminal thresholds. An optional LLM supplies typed proposals; the scientific controller computes likelihoods, updates, costs, and acceptance (Appendix C).

5 EXPERIMENTAL SETUP

Data sources. Molecular and spatial observations come from the Allen whole-mouse-brain atlas (Yao et al., 2023), the whole-mouse-brain MERFISH atlas (Zhang et al., 2023), and the adult human brain single-nucleus atlas (Siletti et al., 2023). Patch-seq supplies paired transcriptomic, morphological, and electrophysiological measurements (Gouwens et al., 2020); public processed

SEA-AD releases supply AD-associated cell states and neuropathology (Gabitto et al., 2024). Registered PK/PD and dynamical models supply mechanistic teacher trajectories (Appendix A.2). Their catalog contains 222,336 A/B train-eligible records (111,168 per grade), 13,945 regimens/target programs, and 49 rejected records. These are eligibility counts before training-split assignment.

Benchmark and splits. The reference graph contains six regions and six compartments, with 15 state channels at 28 times over 672 hours. Directed transport edges drive the occupancy compiler; their symmetrized Laplacian supplies operator context. Holdouts separately exclude datasets/publications, regimens, target programs, and simulator families; repeated measurements from one biological unit stay grouped. Public multi-arm outcomes and teacher replay form separate evaluation strata. The AD-first evaluation uses 64 paired scenario blocks and 2,000 block-bootstrap resamples (seed 3407). Model evaluation specifies five training seeds and independent source-group resampling.

Matched comparisons. Dynamics baselines receive common occupancy inputs and predictive-sample budgets. Acquisition policies share the posterior, feasible query library, outcome channel, and scientific cost units; their objectives and numerical optimizers vary. The main tables identify the comparator families, with adapters and compute caps in Appendix A.9. TRI thresholds are fixed on separate calibration groups and remain unchanged under shift. Failed calls retain their cost and decision loss. Architecture and training details are in Appendices A.3 and C.4.

The evaluation follows four questions, each paired with evidence in the main text:

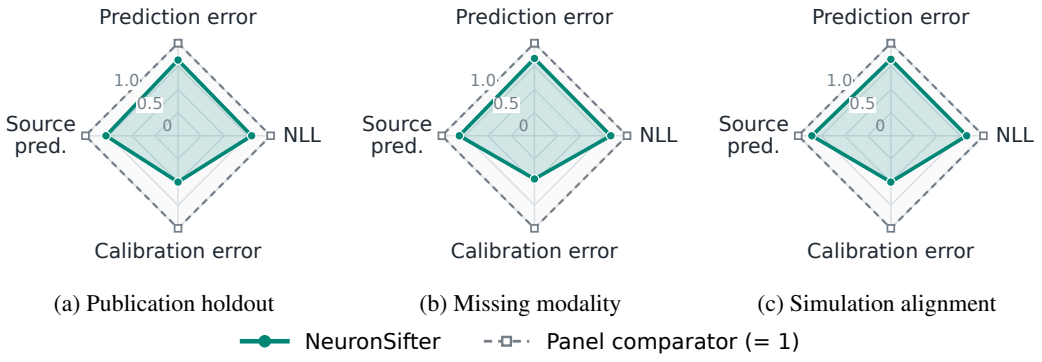
- **RQ1: State inference.** How accurately does the shared posterior transfer across sources and missing modalities? (Figure 3)
- **RQ2: Intervention prediction.** How well does the WAM generalize to held-out regimens and intervention families? (Table 1)
- **RQ3: Evidence acquisition.** Which acquisition strategies reduce intervention risk within a fixed budget? (Table 2)
- **RQ4: Decision reliability.** How do selective acceptance and the framework's components affect decision risk? (Table 3 and Figure 4)

6 Results and Discussion

State inference (RQ1). The shared posterior improves prediction and uncertainty calibration across all three settings in Figure 3. With entire publications held out, prediction error is 0.148 versus 0.181 for scGPT, an 18.2% reduction. With one modality removed, error is 0.182 versus 0.218 for scVI, a 16.5% reduction. Calibration error falls from 0.050 to 0.025 and from 0.075 to 0.035, respectively. Lower NLL in both settings shows that the improvement extends to the predictive distributions. Simulator–observation alignment follows the same pattern: error decreases from 0.195 to 0.161 and calibration error from 0.060 to 0.030. Source predictability also decreases in every setting, consistent with weaker source-specific signals. The radar panels summarize these complementary changes on a common relative scale. Together, improved accuracy and calibration support using the posterior to represent the biological uncertainty carried into intervention comparison and measurement valuation.

Intervention prediction (RQ2). Table 1 compares five external dynamics models under common occupancy conditioning and a 64-sample predictive budget. NeuronSifter achieves held-regimen CRPS 0.110, compared with 0.117 for WDNO and 0.119 for Transolver. On held-out intervention families, CRPS is 0.121 versus 0.130 for WDNO, with paired excess error 0.009 $[0.005, 0.013]$. The corresponding family NRMSE is 0.162 versus 0.176, while the absolute 90% coverage error is 0.006 versus 0.025. Thus the forecast advantage appears in both distributional accuracy and uncertainty calibration under unseen intervention programs. Separate action-encoding controls assess the occupancy compiler's contribution (Table 12); they complement these comparisons between architectures receiving the same inputs.

Evidence acquisition (RQ3). At scientific cost budget 5, NeuronSifter attains terminal risk 0.160, compared with 0.193 for EIG and 0.205 for BALD (Table 2). Decision-aware alternatives are closer: matched numerical BED attains 0.166 and amortized decision-aware BED 0.171. The paired 0.006



(a) Publication holdout (b) Missing modality (c) Simulation alignment

Figure 3: **State inference under source shift and incomplete observations.** Each radius is the metric divided by its comparator value; the comparator equals one and smaller is better. Panels compare with **(a)** scGPT (Cui et al., 2024), **(b)** scVI (Lopez et al., 2018), and **(c)** the alignment control (Gretton et al., 2012). "Source pred." denotes source predictability. All four axes use the same zero-origin scale.

Table 1: External dynamics baselines grouped by modeling approach. All receive common occupancy inputs and 64 predictive samples. Lower is better; gray bold marks best values. NRMSE measures trajectories; Δ_{90} is the absolute error of 90% interval coverage. Full splits and separate action-encoding/ranking controls are in Tables 11, 12 and 31.

	Held regimen	Held family		
Model	CRPS ↓	CRPS ↓	NRMSE ↓	Δ_{90} ↓
NeuronSifter (ours)	**0.110**	**0.121**	**0.162**	**0.006**
Direct neural operators				
Transolver (Wu et al., 2024)	0.119	0.134	0.179	0.088
CoDA-NO (Rahman et al., 2024)	0.125	0.139	0.185	0.103
Mamba-NO (Cheng et al., 2024b)	0.129	0.145	0.192	0.119
Diffusion-based operators				
DiffusionPDE (Huang et al., 2024b)	0.123	0.136	0.184	0.041
WDNO (Hu et al., 2025)	0.117	0.130	0.176	0.025

$[0.003, 0.009]$ gap to numerical BED measures finite-budget planning with the same posterior and decision objective. At the common risk target of 0.22, NeuronSifter requires interpolated cost 2.286, versus 2.375 for numerical BED and 3.410 for EIG. These results connect measurement selection to the downstream intervention decision: gathering information about the latent state and reducing decision loss yield different acquisition priorities. The table groups policies by objective; the budget-allocation comparison is detailed in Appendix A.9.2.

Decision reliability (RQ4). Selective acceptance improves the tradeoff between retaining recommendations and controlling their risk (Table 3). The TRI posterior-risk rule achieves AURC 0.124 versus 0.132 for conformal risk control. At risk 0.20, its accepted fraction is 0.882 versus 0.844, with false acceptance 0.018 versus 0.020. Removing the support gate raises false acceptance to 0.039. These comparisons show how the acceptance rule uses both posterior risk and the available evidence to determine which recommendations to retain.

Figure 4 explains the complementary roles of state uncertainty and evidence acquisition. Fixed-budget risk rises from 0.160 to 0.220 without posterior uncertainty and to 0.248 without evidence acquisition. Removing the terminal gate raises false acceptance from 0.018 to 0.050. The two endpoints therefore distinguish the quality of intervention selection from the reliability of accepting that selection. Configuration mappings and additional controls are in Appendix A.6.1.

Table 2: Acquisition strategies grouped by objective at cost budget 5. $\Delta R = R_{\text{comparator}} - R_{\text{ours}}$ with paired 95% intervals; cost is interpolated at risk 0.22. † marks internal controls. Published methods use task adapters; protocols and compute caps are in Appendix A.9.

Policy	Risk ↓	ΔR [95% CI]	Cost ↓
NeuronSifter (ours)	**0.160**	0	**2.286**
Decision-aware planning			
Matched numerical BED†	0.166	0.006 [0.003, 0.009]	2.375
BAMCP (Guez et al., 2012)	0.169	0.009 [0.000, 0.018]	2.410
Amortized decision-aware BED (Huang et al., 2024a)	0.171	0.011 [0.007, 0.015]	2.500
Earlier decision-focused control†	0.178	0.018 [0.013, 0.023]	2.871
Algorithm- and optimization-directed acquisition			
BAX (Neiswanger et al., 2021)	0.176	0.016 [0.007, 0.025]	2.690
PS-BAX (Cheng et al., 2024a)	0.180	0.020 [0.011, 0.029]	2.805
MF-MES (Takeno et al., 2020)	0.188	0.028 [0.019, 0.037]	3.140
Cost-aware BO (Xie et al., 2024)	0.181	0.021 [0.012, 0.030]	2.540
Information-gain experimental design			
EIG (Lindley, 1956)	0.193	0.033 [0.024, 0.042]	3.410
BALD (Houlsby et al., 2011)	0.205	0.045 [0.036, 0.054]	4.070
DAD (Foster et al., 2021)	0.190	0.030 [0.021, 0.039]	3.280
iDAD (Ivanova et al., 2021)	0.186	0.026 [0.017, 0.035]	3.020
Step-DAD (Hedman et al., 2025)	0.179	0.019 [0.010, 0.028]	2.765
Fixed acquisition heuristics			
Random acquisition	0.240	0.080 [0.071, 0.089]	> 5
Cheapest feasible query	0.229	0.069 [0.060, 0.078]	> 5
Highest-fidelity feasible query	0.208	0.048 [0.039, 0.057]	4.280
Fixed cascade	0.201	0.041 [0.032, 0.050]	3.920

Table 3: **Selective decision reliability.** Accepted fraction is interpolated at risk 0.20; AURC integrates the risk–coverage curve. Full metrics are in Table 20.

Acceptance rule	AURC ↓	Accepted fraction ↑	False acceptance ↓
TRI + posterior risk (ours)	**0.124**	**0.882**	**0.018**
Conformal risk control (Angelopoulos et al., 2024)	0.132	0.844	0.020
Risk without support gate	0.154	0.750	0.039
Predictive uncertainty	0.175	0.650	0.035
Maximum score threshold	0.219	0.473	0.054

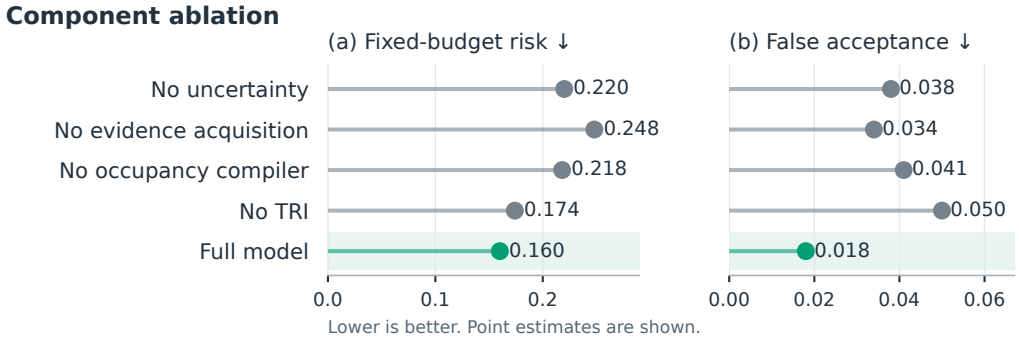


Figure 4: **Components supporting reliable intervention decisions.** Lower is better for fixed-budget risk and false acceptance. The full model is highlighted in green. *No TRI* removes the terminal selective gate; the other controls remove the indicated component. Configuration mappings and numerical values are in Appendix A.6.1.

Additional decision analyses. Intervention-ranking controls quantify the effects of action encoding and posterior dependence (Tables 12 and 19). Independently generated environments test decision quality under transition and observation mismatch (Table 38); frozen-threshold shift tests and LLM backend comparisons are reported in Figure 13 and Appendix A.4.4.

6.1 Agent-in-Twin loop case study

Figure 5 shows how evidence resolves intervention ambiguity in an AD microenvironment. The case compares two programs: A, a high-occupancy anti-amyloid intervention, and B, a moderate-occupancy clearance-enhancing intervention. Initially, high amyloid burden is compatible with uncertain occupancy and clearance capacity; both programs reduce the reference loss, with expected losses 0.275 and 0.283, respectively. The agent first compares occupancy measurement, a clearance biomarker, and a joint panel by prospective net decision value: 0.0153, 0.0313, and 0.0334. It selects the joint panel before either result is known. In the displayed branch, the occupancy readout supports target engagement, while the clearance biomarker supports limited clearance. Assimilation shifts the corresponding marginal probabilities from 0.50 to 0.90. Occupancy alone retains A; the complete evidence reverses the ranking to B, with losses 0.288 for A and 0.187 for B. A post-intervention observation then updates the twin for the next episode. Comparing both programs under this same updated belief gives a loss difference of 0.1005, isolating the value of revising the intervention. The agent evaluates what to measure before deciding what to do (Table 16). The case reports posterior risks and intervention ranking; final acceptance requires the terminal reliability gate. Likelihoods, alternative branches, and the numerical ledger are in Appendix A.5 and Table 15; published AD endpoints remain in Appendix A.4.3.

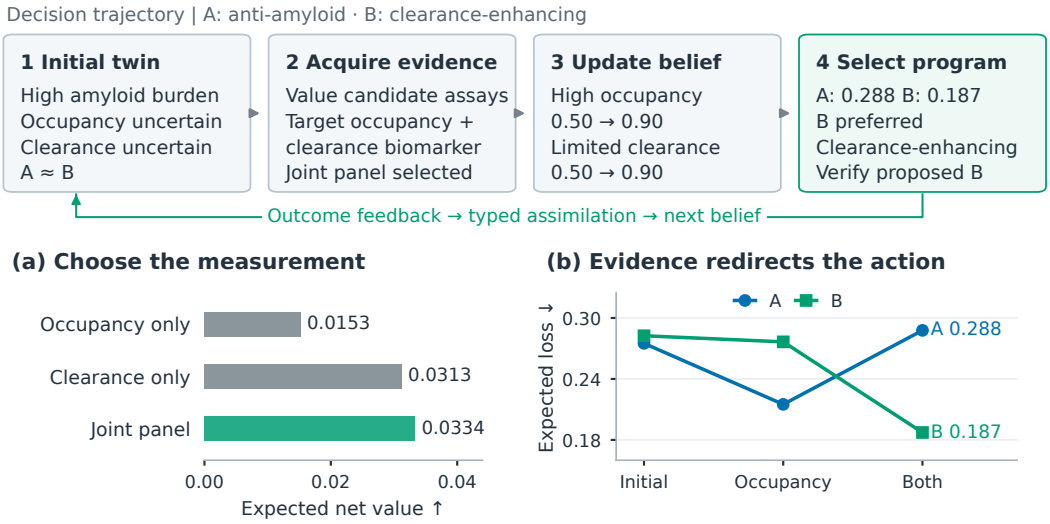


Figure 5: **Agent-in-Twin decision loop. (a)** Prospective net values favor the joint panel. **(b)** Readout assimilation changes the preferred program from A to B. Outcome feedback updates the next belief; final acceptance uses the terminal reliability gate.

7 Conclusion

NeuronSifter introduces a decision-centric *Agent-in-Twin* framework for adaptive intervention discovery in neuronal microenvironments. By coupling a revisable biological digital twin with a scientific agent, NeuronSifter integrates mechanistic state inference, occupancy-based intervention representation, World Action Model prediction, and evidence-driven decision making within a unified loop. The joint biological posterior enables the system to compare competing mechanistic explanations, identify informative observations, and refine intervention strategies as new evidence becomes available. This framework shifts biological AI systems from predicting intervention outcomes toward actively determining which actions, measurements, and model updates are required to improve scientific decisions.

Ethics Statement

NeuronSifter supports retrospective, non-clinical intervention prioritization with public scientific evidence and registered computational tools. Controlled-access participant data follow separate governance and authorization. Registered tool interfaces retain provenance, version, and cost records. Laboratory and clinical use requires domain validation, human authorization, and appropriate oversight; patient-specific treatment additionally requires prospective clinical validation. Publication, demographic, and measurement biases are tracked through source-support and OOD analyses.

Reproducibility Statement

The appendix defines the operator, action compiler, admission rules, grouped splits, baseline interfaces, metrics, and statistical units. Proofs accompany the formal results. The numerical source record includes scenario values, exact Bayesian calculations, and paired block-bootstrap intervals. Figure sources and checks preserve the mapping from each value to its display. Empirical run records include effective configuration, split, checkpoint, tool and environment identities, predictions, costs, failures, and uncertainty estimates.

Published AD trial estimates retain their primary citations and original confidence intervals. Model-to-trial concordance requires frozen replay on the linked endpoints.

Statement on the Use of AI Assistance

AI tools assisted literature retrieval and manuscript organization. The authors retain responsibility for source verification, scientific interpretation, and the final manuscript.

References

P. S. Aisen, S. Gauthier, S. H. Ferris, et al. Tramiprosate in mild-to-moderate Alzheimer's disease - a randomized, double-blind, placebo-controlled, multi-centre study (the Alphase Study). *Archives of medical science : AMS*, 7(1), 2011. doi: 10.5114/aoms.2011.20612.

Daniel Andrews, Shirin Golchi, D. Louis Collins, and Alzheimer's Disease Neuroimaging Initiative. A digital twin methodology using retrospective patient data for sample size reduction in Alzheimer's disease clinical trials. *Alzheimer's Research & Therapy*, 18:145, 2026. doi: 10.1186/s13195-026-02024-5.

Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *International Conference on Learning Representations*, 2024.

Nuutti Barron, Heng Rao, Urmi Saha, Yu Gu, Zhenghao Liu, Ge Yu, Defu Yang, Ashish Raj, and Minghan Chen. Tau-BNO: Brain neural operator for tau transport model. arXiv:2603.08108, 2026.

Danylo Boiko and Viktoriia Mishkurova. Treatment-conditioned diffusion for forecasting neurodegenerative disease progression. arXiv:2605.29932, 2026.

Samantha Budd Haeberlein, P. S. Aisen, F. Barkhof, et al. Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. *The journal of prevention of Alzheimer's disease*, 9(2), 2022. doi: 10.14283/jpad.2022.30.

Charlotte Bunne, Stefan G. Stark, Gabriele Gut, Jacobo S. Del Castillo, Michael Levesque, Kjong-Van Lehmann, Andreas Krause, and Gunnar Ratsch. Learning single-cell perturbation responses using neural optimal transport. *Nature Methods*, 20:1759–1768, 2023. doi: 10.1038/s41592-023-01969-x.

Chu Xin Cheng, Raul Astudillo, Thomas Desautels, and Yisong Yue. Practical bayesian algorithm execution via posterior sampling. In *Advances in Neural Information Processing Systems*, volume 37, 2024a.

Chun-Wun Cheng, Jiahao Huang, Yi Zhang, Guang Yang, Carola-Bibiane Schönlieb, and Angelica I. Aviles-Rivero. Mamba neural operator: Who wins? transformers vs. state-space models for pdes. *arXiv preprint arXiv:2410.02113*, 2024b. URL https://arxiv.org/abs/2410.02113.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, 2014. doi: 10.3115/v1/D14-1179.

Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, Bo Wang, et al. scGPT: Toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, 21:1470–1480, 2024. doi: 10.1038/s41592-024-02201-0.

Michael F. Egan, James Kost, Pierre N. Tariot, et al. Randomized Trial of Verubecestat for Mild-to-Moderate Alzheimer's Disease. *The New England journal of medicine*, 378(18), 2018. doi: 10.1056/nejmoa1706441.

Adam Foster, Desi R. Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3384–3395, 2021.

Mariano I. Gabitto, Kyle J. Travaglini, Victoria M. Rachleff, et al. Integrated multimodal cell atlas of alzheimer's disease. *Nature Neuroscience*, 27:2366–2383, 2024. doi: 10.1038/s41593-024-01774-5.

Yonatan Geifman and Ran El-Yaniv. SelectiveNet: A deep neural network with an integrated reject option. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2151–2159, 2019.

Nathan W. Gouwens, Staci A. Sorensen, Fahimeh Baftizadeh, et al. Integrated morphoelectric and transcriptomic classification of cortical gabaergic cells. *Cell*, 183(4):935–953.e19, 2020. doi: 10.1016/j.cell.2020.09.057.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

Arthur Guez, David Silver, and Peter Dayan. Efficient bayes-adaptive reinforcement learning using sample-based search. In *Advances in Neural Information Processing Systems*, volume 25, 2012.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2555–2565, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640:647–653, 2025. doi: 10.1038/s41586-025-08744-2.

Minsheng Hao, Jing Gong, Xin Zeng, Chiming Liu, Yucheng Guo, Xingyi Cheng, Taifeng Wang, Jianzhu Ma, Xuegong Zhang, and Le Song. Large-scale foundation model on single-cell transcriptomics. *Nature Methods*, 21:1481–1491, 2024. doi: 10.1038/s41592-024-02305-7.

Marcel Hedman, Desi R. Ivanova, Cong Guan, and Tom Rainforth. Step-DAD: Semi-amortized policy-based bayesian experimental design. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 22904–22923. PMLR, 2025.

Rebecca D. Hodge, Trygve E. Bakken, Jeremy A. Miller, et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature*, 573:61–68, 2019. doi: 10.1038/s41586-019-1506-7.

Lawrence S. Honig, Bruno Vellas, Michael Woodward, et al. Trial of Solanezumab for Mild Dementia Due to Alzheimer's Disease. *The New England journal of medicine*, 378(4), 2018. doi: 10.1056/nejmoa1705971.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. arXiv:1112.5745, 2011.

Peiyan Hu, Rui Wang, Xiang Zheng, Tao Zhang, Haodong Feng, Ruiqi Feng, Long Wei, Yue Wang, Zhi-Ming Ma, and Tailin Wu. Wavelet diffusion neural operator. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/20eadb6338a53114a6371f102c5ded95-Abstract-Conference.html.

Daolang Huang, Yujia Guo, Luigi Acerbi, and Samuel Kaski. Amortized bayesian experimental design for decision-making. In *Advances in Neural Information Processing Systems*, volume 37, pp. 109460–109486, 2024a.

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. DiffusionPDE: Generative PDE-solving under partial observation. In *Advances in Neural Information Processing Systems*, volume 37, 2024b. doi: 10.52202/079017-4140. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/eb3878c1dcbfff9ee95d5d033e5f5942-Abstract-Conference.html.

Yinyu Huang, Yilin Zhang, Sofia Michopoulou, Christopher Kipps, and Rahman Attar. Transition-based digital twin modelling for alzheimer's disease under sparse longitudinal data. arXiv:2606.09671, 2026.

Desi R. Ivanova, Adam Foster, Steven Kleinegesse, Michael U. Gutmann, and Tom Rainforth. Implicit deep adaptive design: Policy-based experimental design without likelihoods. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Ian T. Jolliffe and Jorge Cadima. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065):20150202, 2016. doi: 10.1098/rsta.2015.0202.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1–2):99–134, 1998. doi: 10.1016/S0004-3702(98)00023-X.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956. doi: 10.1214/aoms/1177728069.

Wenxi Liu, Michael Trimboli, and Xianqi Li. Ai-augmented adaptive digital twin modeling for brain tumor evolution prediction and treatment scheduling. arXiv:2607.13877, 2026.

Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15:1053–1058, 2018. doi: 10.1038/s41592-018-0229-2.

Mohammad Lotfollahi, Anna Klimovskaia Susmelj, Carlo De Donno, et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular Systems Biology*, 19(6):e11517, 2023. doi: 10.15252/msb.202211517.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George E. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Jiangran Lyu, Ziming Li, Xuesong Shi, Chaoyi Xu, Yizhou Wang, and He Wang. DyWA: Dynamics-adaptive world action model for generalizable non-prehensile manipulation. arXiv:2503.16806, 2025.

MICrONS Consortium. Functional connectomics spanning multiple areas of mouse visual cortex. *Nature*, 640:435–447, 2025. doi: 10.1038/s41586-025-08790-w.

Model Context Protocol Contributors. Model Context Protocol Specification: Tools. Protocol revision 2025-06-18, 2025. URL https://modelcontextprotocol.io/specification/2025-06-18/server/tools. Accessed September 21, 2026.

Arnau Montagud. Systems biology at the body scale: Toward adaptive digital twins in precision medicine. *Current Opinion in Systems Biology*, 44:100592, 2026. doi: 10.1016/j.coisb.2026.100592.

Willie Neiswanger, Ke Alexander Wang, and Stefano Ermon. Bayesian algorithm execution: Estimating computable properties of black-box functions using mutual information. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8005–8015, 2021.

Aida Nonn, Elma Kerz, Daniel Wiechmann, and Paul Edison. Towards a physics-informed multiscale digital twin for precision medicine in Alzheimer's disease. arXiv:2608.15342, 2026.

Andrzej Przekwas, Carly Norris, and Harsha T. Garimella. Mechanistic modeling of amyloid dynamics relating to alzheimer's disease progression. *Frontiers in Aging Neuroscience*, 18:1730480, 2026. doi: 10.3389/fnagi.2026.1730480.

Lu Qiu, Yizhuo Li, Yi Chen, Yuying Ge, Yixiao Ge, and Xihui Liu. Making foresight actionable: Repurposing representation alignment in world action models. arXiv:2606.12217, 2026.

Md Ashiqur Rahman, Robert Joseph George, Mogab Elleithy, Daniel Leibovici, Zongyi Li, Boris Bonev, Colin White, Julius Berner, Raymond A. Yeh, Jean Kossaifi, Kamyar Azizzadenesheli, and Anima Anandkumar. Pretraining codomain attention neural operators for solving multiphysics pdes. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/bc75fa9843a7905bbed9d83895a88f7f-Abstract-Conference.html.

Balaraman Ravindran and Andrew G. Barto. Approximate homomorphisms: A framework for non-exact minimization in markov decision processes. In *Proceedings of the Fifth International Conference on Knowledge Based Computer Systems*, 2004.

Yusuf Roohani, Kexin Huang, and Jure Leskovec. Predicting transcriptional outcomes of novel multigene perturbations with GEARS. *Nature Biotechnology*, 42:927–935, 2024. doi: 10.1038/s41587-023-01905-6.

Yanay Rosen, Yusuf Roohani, Ayush Agrawal, Leon Samotorčan, Tabula Sapiens Consortium, Stephen R. Quake, Jure Leskovec, et al. Universal cell embedding provides a foundation model for cell biology. *Nature*, 2026. doi: 10.1038/s41586-026-10689-z.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/eb0b13cc515724ab8015bc978fdde0ad-Abstract-Conference.html.

Seattle Alzheimer's Disease Brain Cell Atlas Consortium. SEA-AD data and downloads. Allen Institute data documentation, 2026. URL https://brain-map.org/consortia/sea-ad/our-data.

Andrew N. Shen, Katelin S. Matazel, W. Drew Gill, Lorna Ewart, Randy S. Daughters, and Hector Rosas-Hernandez. Modeling neurovascular dysfunction in alzheimer's disease using an isogenic brain-chip model. *Fluids and Barriers of the CNS*, 23:1, 2026. doi: 10.1186/s12987-025-00708-y.

Hailing Shi, Yichun He, Yiming Zhou, et al. Spatial atlas of the mouse central nervous system at molecular resolution. *Nature*, 622:552–561, 2023. doi: 10.1038/s41586-023-06569-5.

Kimberly Siletti, Rebecca Hodge, Alejandro Mossi Albiach, et al. Transcriptomic diversity of cell types across the adult human brain. *Science*, 382(6667):eadd7046, 2023. doi: 10.1126/science.add7046.

John R. Sims, Jennifer A. Zimmer, Cynthia D. Evans, et al. Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA*, 330(6), 2023. doi: 10.1001/jama.2023.13239.

Leon Stefanovski, Paul Triebkorn, Andreas Spiegler, Margarita-Arimatea Diaz-Cortes, Ana Solodkin, Viktor Jirsa, Anthony Randal McIntosh, Petra Ritter, and Alzheimer's Disease Neuroimaging Initiative. Linking molecular pathways and large-scale computational modeling to assess candidate disease mechanisms and pharmacodynamics in alzheimer's disease. *Frontiers in Computational Neuroscience*, 13:54, 2019. doi: 10.3389/fncom.2019.00054.

Shion Takeno, Hitoshi Fukuoka, Yuhki Tsukada, Toshiyuki Koyama, Motoki Shiga, Ichiro Takeuchi, and Masayuki Karasuyama. Multi-fidelity bayesian optimization with max-value entropy search and its parallelization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9334–9345, 2020.

Christina V. Theodoris, Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, Helen Mantineo, Emily M. Brydon, Zexian Zeng, X. Shirley Liu, and Patrick T. Ellinor. Transfer learning enables predictions in network biology. *Nature*, 618:616–624, 2023. doi: 10.1038/s41586-023-06139-9.

Christopher H. van Dyck, Chad J. Swanson, Paul Aisen, et al. Lecanemab in Early Alzheimer's Disease. *The New England journal of medicine*, 388(1), 2023. doi: 10.1056/nejmoa2212948.

Jindong Wang, Yutong Mao, Xiao Liu, Wenrui Hao, and Alzheimer's Disease Neuroimaging Initiative. Learning patient-specific spatial biomarker dynamics via operator learning for Alzheimer's disease progression. *npj Systems Biology and Applications*, 12:104, 2026. doi: 10.1038/s41540-026-00719-x. Preprint first released in 2025 as arXiv:2507.16148.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53681–53705, 2024. URL `https://proceedings.mlr.press/v235/wu24r.html`.

Qian Xie, Raul Astudillo, Peter I. Frazier, Ziv Scully, and Alexander Terenin. Cost-aware bayesian optimization via the pandora's box gittins index. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3669.

Zizhen Yao, Cindy T. J. van Velthoven, Michael Kunst, et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature*, 624:317–332, 2023. doi: 10.1038/s41586-023-06812-z.

Seonghyeon Ye, Yunhao Ge, Kaiyuan Zheng, Shenyuan Gao, Sihyun Yu, et al. World action models are zero-shot policies. arXiv:2602.15922, 2026.

Meng Zhang, Xingjie Pan, Won Jung, et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature*, 624:343–354, 2023. doi: 10.1038/s41586-023-06808-9.

Appendix navigation

Experiments

A Supplementary experiments 15
A.1 CNS experiment protocol 15
A.2 Dataset construction and preprocessing 19
A.3 Hyperparameters and model scales 20
A.4 State and dynamics results 20
A.5 Agent-in-Twin decision trajectory 25
A.6 Robustness and mechanism tests 27
A.7 Statistical analysis 30
A.8 Compute, cost, and reproducibility 31
A.9 Baseline coverage and independent environments 32

Method, theory, and interpretation

B Notation and assumptions 36
C Extended Agent-in-Twin method 37
D Theoretical results 40
E Proofs 43

A Supplementary Experiments

A.1 CNS Experiment Protocol

A.1.1 Tasks, units, and endpoints

The evaluation unit is a microenvironment context with initial evidence, supported intervention programs, and declared outcomes. Atlases assess observation heads; mechanistic teachers assess controlled dynamics; independent biological outcomes assess intervention response. Their evidence strata remain separate.

Continuous predictions use training-only centering/scaling. For predictive law F, $\mathrm{CRPS}(F, y) = \mathbb{E}|X - y| - \frac{1}{2}\mathbb{E}|X - X'|$ with independent draws; NRMSE is root mean squared error in the declared normalized coordinates. Average observed channels and times within each episode before averaging independent episodes, retaining the same missingness mask across methods. Three distinct quantities are named separately throughout: *accepted fraction* is the share of contexts a selective rule accepts, *bound coverage* is the observed coverage of the frozen upper-loss bound at its nominal level, and *calibration error* is the absolute deviation of a predictive interval's observed coverage from its nominal level. Spearman correlation assesses ranks; ordering accuracy assigns half credit to ties. Intervention regret is $\ell(\widehat{a}) - \min_{a \in \mathcal{A}_i} \ell(a)$ within the supported set. Query regret compares net utility with the best feasible query. Threshold costs use linear interpolation; AURC uses trapezoidal integration with the limiting risk at zero coverage. Fractional threshold costs are summaries, not executable fractional queries.

A.1.2 State, dynamics, and dependence controls

Whole-source holdouts and modality removal assess state transfer, joint likelihood, coverage, and cross-modal consistency against unimodal and late-fusion controls. Trainable methods use five seeds with matched optimization and inference budgets. Only QC-approved teacher records enter training, with teacher-uncertainty weighting. Regimen, target-program, horizon, and simulator-family holdouts exercise the complete stochastic operator and all mechanism experts.

Action-ID, scalar-exposure, flattened-field, parameter-matched shared-operator, and parallel-expert variants are internal ablations. Identical prefixes with different future programs test past-state invariance; matched integrated exposure/occupancy with different timing tests scalar information loss; matched programs under different transport/inflammation contexts test state-dependent response. Matching variables and cutoffs are fixed before outcome access. Joint-posterior, point-state, and dependence-ablated controls share marginal calibration and physical validity. Pairing Monte Carlo seeds improves precision; it does not define biological dependence.

A.1.3 Loss preferences and sensitivity

The amyloid, tau, inflammation, and circuit-dysfunction terms in Equation (7) represent distinct pathological and functional domains. Clinical benefit, failed amyloid interventions, and mixed trial findings motivate retaining those distinctions (van Dyck et al., 2023; Egan et al., 2018; Honig et al.,

Table 4: CNS evaluation tasks, comparisons, holdouts, and endpoints.

Task	Comparison	Holdout	Endpoint
State inference	Joint versus unimodal/late-fusion inference	Whole source and missing modality	Native readout likelihood and calibration
Intervention dynamics	Occupancy versus identity/scalar exposure; ordered versus parallel experts	Regimen, target program, horizon, simulator	Trajectory CRPS, coverage, physical constraints
Intervention ranking	Joint, point-state, and dependence-ablated posteriors	Initial context and supported action family	Regret and ordering on matched functional outcomes
Evidence planning	Decision-directed versus uncertainty/random queries; full-posterior external equivalence	Biological mechanism and measurement context	Terminal risk at cost; cost at target risk
Reliability	TRI versus matched selective controls	Source, regimen, and measurement shift	Risk–coverage, false acceptance, utility and cost

Table 5: Baseline families and matched interfaces. Citations identify published methods or comparator ancestry; internal ablations and reference policies are defined in the protocol.

Task	Baseline families	Matched inputs and purpose
State	PCA (Jolliffe & Cadima, 2016), scVI (Lopez et al., 2018), Geneformer (Theodoris et al., 2023), scGPT (Cui et al., 2024), UCE (Rosen et al., 2026), scFoundation (Hao et al., 2024); linear, from-scratch and single-source controls	Frozen sources, splits, feature mapping and probe/fine-tune budget; separate transfer from source memorization.
Dynamics	GRU (Cho et al., 2014), CPA (Lotfollahi et al., 2023), CellOT (Bunne et al., 2023), GEARS (Roohani et al., 2024), PlaNet (Hafner et al., 2019), Dreamer (Hafner et al., 2020), Transolver (Wu et al., 2024), WDNO (Hu et al., 2025), DiffusionPDE (Huang et al., 2024b), CoDA-NO (Rahman et al., 2024), Mamba-NO (Cheng et al., 2024b); persistence, linear, action-ID and scalar-exposure controls	Common observations, actions, horizon, optimization and predictive samples; isolate structured action and stochastic dynamics.
Acquisition	EIG (Lindley, 1956), BALD (Houlsby et al., 2011), BAX (Neiswanger et al., 2021), PS-BAX (Cheng et al., 2024a), MF-MES (Takeno et al., 2020), DAD (Foster et al., 2021), iDAD (Ivanova et al., 2021), Step-DAD (Hedman et al., 2025), decision-aware BED (Huang et al., 2024a), cost-aware BO (Xie et al., 2024), BAMCP (Guez et al., 2012); random, cheapest/highest-fidelity and fixed-cascade controls	Common candidates, retained outcomes, availability masks, cost and stopping; objectives and adapters in Appendix A.9; all families have entries in Tables 2, 14 and 26.
Selection	Conformal risk control (Angelopoulos et al., 2024), SelectiveNet (Geifman & El-Yaniv, 2019); score, uncertainty, calibration-only and OOD-only controls	Common calibration groups, coverage grid and test thresholds; test support-aware acceptance.
Scale	Student, Large and XL profiles; no-distillation Student, fixed threshold and fixed route	Common candidate set, decision rule and end-to-end cost; test adaptive escalation.

2018; Budd Haeberlein et al., 2022); these outcomes do not estimate the weights or calibrate the endpoint bridge.

The proposed empirical reference scales harmful channels monotonically to $[0, 1]$ using training-reference 1st/99th percentiles, recording clipped fractions. Constant or unsupported channels cannot define a calibrated loss. Normalize time weights and region weights separately to sum to one. Equal endpoint weights $(1/4, 1/4, 1/4, 1/4)$, uniform region weights, and $\lambda_D = 0.1$ for $[0, 1]$-scaled intervention burden define transparent reference preferences, not biological constants or clinically elicited utility. Configure them before evaluation; hard support, exposure, and safety constraints remain separate.

Prespecify one-at-a-time weight multipliers $\{0.5, 1, 2\}$ with renormalization, endpoint-omission diagnostics, and a simplex uncertainty set. Vary $\lambda_D \in \{0, 0.05, 0.1, 0.2\}$ and repeat with 5th/95th-percentile scaling. Evaluate every policy under each shared setting; report top-action agreement, Kendall correlation, paired excess loss, acceptance coverage, and worst preference-dependent regret. Retain sign reversals and unsupported regions. Table 6 reports these endpoints on the evaluation scenarios: the selected action is retained on 0.797–0.953 of contexts, paired excess loss stays at or below 0.016, and the simplex uncertainty set is the binding case. These preference-sensitivity results do not establish a clinical minimum important difference.

Table 6: **Preference robustness of the intervention ranking.** Prespecified perturbations of Equation (7) under a shared posterior, query library, and stopping rule. Agreement, correlation, and accepted fraction are higher-better; excess loss and worst-case regret are lower-better.

Preference setting	Top-action agreement ↑	Kendall corr. ↑	Paired excess loss ↓	Accepted fraction ↑	Worst-case regret ↓
Reference preferences	1.000	1.000	0.000	0.882	0.012
Endpoint weight $\times 0.5$	0.938	0.902	0.004	0.871	0.017
Endpoint weight $\times 2$	0.906	0.869	0.006	0.863	0.021
Endpoint omission	0.844	0.795	0.011	0.842	0.029
Intervention-burden weight 0 or 0.2	0.922	0.884	0.005	0.868	0.019
Percentile rescaling 5/95	0.953	0.925	0.003	0.875	0.015
Simplex uncertainty set	0.797	0.742	0.016	0.828	0.036

Table 7: Biological measurements and the uncertainty they address.

Readout	Posterior information	Decision role
CNS exposure / transport	Free concentration and barrier transport	Dose, route, and schedule
Target engagement	Occupancy and binding	Engagement and target program
Protein burden / clearance	Amyloid, tau, and turnover	Production versus clearance programs
Inflammatory response	Cell-state and mediator response	Inflammatory benefit and functional cost
Electrophysiology / structure	Excitability, instability, morphology	Function beyond molecular response

A.1.4 Measurement menu and feedback

Each query requires an admissible outcome source, observation model, and update scope. Plasma/CSF assays require an explicit target-site observation mapping. Invasive measurements are hybrid actions with a state transition and cost. Complete outcome matrices or mechanistic replay supply common feasible query outcomes; logged evaluation requires policy overlap. Partial outcomes update admitted fields, while failures retain status and incurred cost. Fixed-budget episodes compare terminal loss; fixed-risk episodes compare cost. Prespecified groups and failures remain in the denominator.

A.1.5 Matched non-amortized decision-aware planner

Matched numerical BED shares NeuronSifter's posterior, compiler, transition, retained-outcome channel, query library, terminal loss, costs, maximum lookahead, stopping options, and simulation budget. The existing manuscript reports risk 0.166 versus 0.160 and costs 2.375 versus 2.286 to risk 0.22. These are finite-budget planner comparisons. Running the same full-posterior policy through an external interface gives the identical result (Table 14); changing an interface alone supplies no additional information or attainable value.

The revised comparison specifies uniform nested-MC allocation for BED and adaptive allocation of the same estimator for NeuronSifter, with fixed pilot rules, depth, call accounting, and stopping in Appendix A.9.2. The complete baseline inventory, computational budgets, and budget-sensitivity results are in Appendix A.9. Paired allocation runs isolate the source of the gap; architecture effects are evaluated separately in Appendix A.9.4.

A.1.6 LLM policy-backend comparison

Kimi K3, DeepSeek V4 Flash, and GPT-5.6-Sol propose typed actions or stopping through the same frozen twin. Match posterior summaries, feasible libraries, prompts, exemplars, evidence order, context selection, rollout budget, updates, and limits on rounds, attempts, retries, and wall time. Paired exogenous randomness supplies the conditional outcome channel; memory resets and evaluation targets remain hidden. Provider/model version, date, reasoning/decoding settings, context limits, seed support, tokens, billing, and latency accompany executions. Unavailable backends remain missing.

Terminal risk includes every episode under a common fallback rule. Invalid proposals are divided by all returned proposals including retries; no-return episodes retain refusal/timeout status and an undefined proposal rate. Executed calls exclude rejected proposals, while cost and deadline-capped time include attempts. Average five repetitions per episode/backend before source-group bootstrap. The three risk contrasts form one Holm-corrected family, with no test-based backend selection.

A.1.7 AGENT–TWIN PROMPTS AND MCP EXCHANGE

These examples specify an optional Model Context Protocol (MCP) adapter using the 2025-06-18 tools interface (Model Context Protocol Contributors, 2025). They are designed messages, not a recorded run or evidence of a deployed MCP server; the numerical policy does not require an LLM. The analytic posterior update below uses the exact exposure example in Appendix A.4.2.

Agent system prompt

```
You plan CNS interventions through the Twin's registered tools.
Compare supported structured actions under the supplied loss and
budget, using one belief ID/version for every candidate.  Proposed
interventions are do-actions, not evidence.
Choose queries by net decision value.  Let the Twin validate evidence
and update complete joint particles.  Never invent outcomes, costs,
likelihoods, or TRI; retain failures and incurred cost.  Missing
support or calibration requires insufficient_evidence and the missing
fields.  Do not fill them from evaluation outcomes or remembered trial
results.
```

Episode request. Use belief_demo, version 0; budget 5 protocol cost units; the Standard, exposure-adapted, and inflammation-sparing library. Return the selected action/query ID, belief version, reason, incurred cost, and missing reliability fields. The observations follow the analytic exposure example.

The adapter binds the posterior to a server-side state handle: the client passes the expected version, and the server validates the registered evidence reference before assimilation. The following is one tool entry from a `tools/list` response, whose nested scientific evidence is resolved through the existing registry rather than supplied as an unconstrained LLM-generated value.

MCP tool definition: registry-backed feedback

```
{"name":"ingest_feedback",
 "description":"Validate registered
     evidence and update a joint posterior
     .",
 "inputSchema":{"type":"object","
     additionalProperties":false,
   "required":["belief_id","
     expected_version","evidence_ref"],
   "properties":{"belief_id":{"type":"
     string"},
     "expected_version":{"type":"integer
     ","minimum":0},
     "evidence_ref":{"type":"string"}}}}
```

Agent request: MCP tools/call

```
{"jsonrpc":"2.0","id":17,"method":"tools/
     call",
 "params":{"name":"ingest_feedback","
     arguments":{
   "belief_id":"belief_demo","
     expected_version":0,
   "evidence_ref":"example://case/
     exposure/low_exposure"}}}
```

Twin response: protocol success with insufficient evidence

```
{"jsonrpc":"2.0","id":17,"result":{
 "content":[{"type":"text","text":"
     Posterior updated; support and TRI
     incomplete."}],
 "structuredContent":{"belief_id":"
     belief_demo","version":1,
   "states":["exposure_limited","
     inflammation_sensitive"],
   "weights":[0.8,0.2],"evidence_type":"
     analytic_example",
   "decision":"insufficient_evidence","
     realized_cost_units":null,
   "missing":["validated_action_support
     ","calibrated_TRI","resource_receipt
     "]},
 "isError":false}}
```

Figure 6: **MCP feedback exchange.** Left: advertised tool and matching request. Right: analytic posterior update with a scientific insufficient-evidence decision. Protocol success and intervention acceptance are separate.

The weights follow $P(E \mid \text{low exposure}) = (0.5 \times 0.8)/(0.5 \times 0.8 + 0.5 \times 0.2) = 0.8$. MCP success confirms tool execution, not intervention acceptance, and every later rollout and query-value computation uses version 1. A stale-version call returns an explicit tool error and requires a fresh state read; malformed evidence does not advance the posterior version. An executed backend record retains prompts, model/decoding settings, schemas, calls, failures, and posterior versions.

Interaction sequence. Read the posterior with `infer_world_state`; compare supported programs through `rollout_counterfactuals`; choose a feasible query or stop through `plan_interventions`; validate and assimilate its registered result through `ingest_feedback`. Recompute candidate losses/query values under the updated version and

apply TRI before a terminal decision. The Twin owns scientific state, admission, and realized-cost accounting throughout.

A.1.8 COMPILER, TEACHER, AND ACQUISITION STRESS TESTS

The finite compiler reference assigns $(0.35, 0.25, 0.20, 0.12, 0.08)$ to low-burden, pathology-dominant, inflammation-dominant, mixed, and high-burden outcomes. Perturbation transfers mass from the second/third bins to the first. Exact TV and forward KL characterize these fixed categories, not full-state fidelity. Physical-teacher bins are fixed on training outcomes; raw and compiled programs share initial state, noise, dose, route, and schedule.

Train on one registered solver family and evaluate independently specified dynamics, compiler parameters, and parameter distributions from a held family or resolution. Retain shared-teacher replay as a consistency control. Shift transport/affinity log-prior means by 0, 0.25, 0.5, and one prior SD; separately omit inflammatory coupling, double observation variance, or multiply state-dependent admission odds by four. QC grades establish numerical provenance; biological teacher accuracy remains unmeasured. Structural clamps use observationally compatible parameter scenarios (Appendix C.10).

EIG uses nested-Monte-Carlo $I(\boldsymbol{\xi}; \widetilde{o} \mid b, q)$; regime-target BALD uses $H[m \mid b] - \mathbb{E}H[m \mid b, q, \widetilde{o}]$. Finite-query iDAD/Step-DAD adapters retain their information objectives, with online refinement for Step-DAD (Ivanova et al., 2021; Hedman et al., 2025). Amortized BED retains the common terminal loss (Huang et al., 2024a). Match posterior, feasible library, channel, stopping, realized costs, and amortized parameter/simulation budgets. Top-query agreement uses 64 common beliefs and fixed tie breaking; decision risk uses each policy's own closed loop.

Surrogate controls. The mechanistic–residual MPC adapter propagates common posterior scenarios and uses the shared decision-value measurement channel (Liu et al., 2026). Tau transport and treatment-history diffusion adapters align common observables, action support, and loss (Barron et al., 2026; Boiko & Mishkurova, 2026). Same-occupancy Transolver replaces diffusion with direct prediction while retaining posterior/occupancy draws, graph, observations, split, horizon, and a declared residual likelihood (Wu et al., 2024). The same-compiler solver uses inferred parameters, not privileged truth. DreamerV3 retains behavior learning under matched histories, typed feasibility, loss, and interaction budget (Hafner et al., 2025); FNO is an operator-family reference where discretization permits (Li et al., 2021). Table 31 report the prescribed accuracy, constraint, and cost endpoints.

Calibration and uncertainty. Freeze policy, stopping, TRI thresholds, and the in-distribution calibrator on disjoint groups. Tool shift doubles measurement variance and halves completion odds; admission shift multiplies state-dependent odds by four. Marginalize rejected values while retaining modeled admission events. Report upper-loss coverage at nominal 0.80/0.85/0.90/0.95 and the full risk–coverage curve without retuning. The 64 paired scenario blocks use centered paired deviations, 2,000 bootstrap resamples, and seed 3407; binary counts use Wilson intervals. Empirical resampling and separate optimization variation follow Appendix A.7. Training, calibration, biological outcomes, and teacher replay retain their distinct groups and roles.

A.2 DATASET CONSTRUCTION AND PREPROCESSING

Public evidence families. Molecular and spatial references include whole-brain mouse atlases (Yao et al., 2023; Zhang et al., 2023; Shi et al., 2023), human and cross-species cell taxonomies (Siletti et al., 2023; Hodge et al., 2019), Patch-seq molecular/morphological/electrophysiological measurements (Gouwens et al., 2020), and MICrONS functional connectomics (MICrONS Consortium, 2025). Disease, intervention, model, and tool sources retain separate licenses and source cards.

AD-first resource roles. SEA-AD links molecular state with neuropathology (Gabitto et al., 2024; Seattle Alzheimer's Disease Brain Cell Atlas Consortium, 2026). This public-only path uses open processed releases and documented summaries. Release identifiers and corrected metadata are fixed before grouping.

Admission, scientific units, and splits. Records retain measurement identity: cross-sectional severity is a state observation; Patch-seq modalities share a cell, while unrelated specimens remain separate. Intervention evaluation records program, dose/schedule, context, time, comparator, and replicate-level outcomes. Sources with partial coverage contribute to the corresponding state or mechanism task.

Source cards preserve release, license, version, provenance, parser/mapping, redistribution, uncertainty, and split role. Admission requires units, scientific identity, schema, and evidence grade. Cell

Table 8: Biological resources and their roles in CNS evaluation.

Resource	Microenvironment information	Evaluation role
Public processed SEA-AD	AD cell states and quantitative neuropathology	Disease-state inference and pathology calibration
Allen molecular/spatial atlases	Cell identity, anatomical location, regional context	Observation alignment and cross-source state transfer
Patch-seq / cell-types data	Linked molecular, electrophysiological, and morphological measurements	Cross-modal state inference and functional consistency
MICrONS	Circuit structure linked to functional observations	Circuit-context modeling within the measured tissue
Neurovascular brain-chip assays	Transporter function and barrier phenotypes	Targeted external mechanistic checks on their donor/model scope
Registered PK/PD and dynamical models	Exposure, occupancy, transport, and biological trajectories	Uncertainty-weighted teachers and mechanistic replay
Admissible multi-arm outcomes	Matched interventions and functional measurements	Biological ranking and acquisition evaluation

Ontology, UBERON, molecular identifiers, and anatomical coordinates standardize records while preserving original values. Subject, sample, cell, assay, intervention-family, simulator, tool, and publication groups constrain splits. Repeated observations inherit groups; source/temporal partitioning precedes preprocessing. Training, validation, calibration, test, and shift sets remain disjoint.

WAM examples and query matrices. WAM examples include initial context, complete programs, compiled fields, masks, targets, and source quality. Teachers retain uncertainty and provenance. Query rows record tool/version, input, fidelity, schema, reserved/realized cost, latency, status, uncertainty, provenance, and admitted fields. Complete matrices support matched replay. Execution status and biological outcomes occupy separate fields. Registry exports supply modality frequencies and independent-unit counts. Disease-specific protocols define extension beyond AD.

A.3 Hyperparameters and Model Scales

The declared native backbone uses width 696, eight diffusion blocks, 12 attention heads, 96 mechanism-state tokens, 36 graph modes, and three temporal wavelet levels. Dropout is 0.05 and the feed-forward multiplier is four. Ten ordered mechanism experts use residual scale 0.08 with a shared coupling scale of 0.04. These values specify the reference architecture; executed settings are recorded with each run.

The experiment protocol varies data fraction $\{10, 25, 50, 100\}\%$, source diversity, posterior sample count, rollout horizon, denoising steps, and acquisition budget. Student, Large, and high-fidelity profiles separate throughput from decision quality. Selected checkpoints, parameter counts, training budgets, FLOPs, and measured memory are supplied by execution records. Execution records also identify the hardware and run associated with each measured result.

A.4 Supplementary State and Dynamics Results

The state, dynamics, and mechanism controls below retain separate holdout strata and measurement scales: state transfer tests source/modality generalization, dynamics tests held regimens/compositions/simulators, and mechanism diagnostics distinguish prefix consistency, occupancy timing, and biological context. Their endpoint definitions and matched masks follow Appendix A.1. State-transfer results appear in the main-text radar panels (Figure 3); the following tables report supplementary dynamics and mechanism controls.

A.4.1 WAM architecture comparison

Table 1 summarizes the five named external baselines; Table 11 adds all seven holdout strata and the two internal operator controls. The eight-operator suite uses identical initial fields, posterior draws, occupancy paths, masks, and QC-approved records. Deterministic NeuronSifter removes trajectory diffusion from the same backbone, the factor-coupled model is the internal deterministic reference, and external methods retain their native architectures and objectives, mapped to the common CNS grid. Training uses seeds 40–44, validation-only selection, and the registered matched budgets. All forecasts cover 28 times and 15 channels under a common 64-sample predictive budget; every model propagates input-posterior uncertainty and stochastic operators additionally sample conditional

Table 9: Dynamics holdouts; calibration error is the absolute deviation from nominal 90 percent.

Split	Action model	NRMSE ↓	CRPS ↓	Calib. err. ↓	Rank corr. ↑	Constraint viol. ↓
Held regimen	Structured WAM (ours)	**0.150**	**0.110**	**0.025**	**0.820**	**0.012**
	action ID	0.215	0.165	0.065	0.690	0.031
Held composition	Structured WAM (ours)	**0.173**	**0.128**	**0.035**	**0.790**	**0.017**
	flattened fields	0.210	0.151	0.059	0.720	0.028
Held simulator	Structured WAM (ours)	**0.190**	**0.145**	**0.046**	**0.750**	**0.022**
	neural operator (Wu et al., 2024)	0.228	0.172	0.075	0.680	0.038

Table 10: Controlled intervention diagnostics. Lower is better. Prefix discrepancy uses normalized state coordinates; timing and context contrasts report terminal decision risk.

Controlled comparison	Occupancy program (ours)	Scalar exposure	Action identity
Prefix consistency	**0.007**	0.018	0.022
Exposure-matched timing	**0.162**	0.214	0.231
Context-dependent response	**0.176**	0.223	0.241

trajectories. Each strict split excludes its named identity. CRPS uses common observed coordinates, while family NRMSE and absolute 90% coverage deviation jointly assess accuracy and uncertainty.

Table 11: **WAM baselines under common occupancy conditioning.** Evaluation results; lower is better. Seven columns report CRPS. The final two report trajectory-family NRMSE and $\Delta_{90} = |\widehat{\text{coverage}} - 0.90|$.

	CRPS across held-out contexts ↓							Family ↓	
Model	Random	Family	Regimen	Route	Target	Simulator	Source	NRMSE	Δ_{90}
NeuronSifter-Large (ours)	**0.095**	**0.121**	**0.110**	**0.117**	**0.134**	**0.145**	**0.151**	**0.162**	**0.006**
Deterministic NeuronSifter	0.105	0.143	0.130	0.138	0.158	0.170	0.181	0.171	0.119
Factor-coupled operator	0.116	0.158	0.142	0.154	0.179	0.191	0.204	0.204	0.166
Transolver (Wu et al., 2024)	0.101	0.134	0.119	0.128	0.149	0.172	0.180	0.179	0.088
DiffusionPDE (Huang et al., 2024b)	0.100	0.136	0.123	0.130	0.148	0.162	0.177	0.184	0.041
WDNO (Hu et al., 2025)	0.099	0.130	0.117	0.124	0.142	0.157	0.168	0.176	0.025
CoDA-NO (Rahman et al., 2024)	0.103	0.139	0.125	0.134	0.153	0.166	0.181	0.185	0.103
Mamba-NO (Cheng et al., 2024b)	0.105	0.145	0.129	0.136	0.158	0.174	0.187	0.192	0.119

In this comparison, family CRPS is 0.121 for NeuronSifter, 0.130 for WDNO, and 0.143 for deterministic NeuronSifter. The WDNO excess is 0.009 (95% paired interval $[0.005, 0.013]$; 64 paired blocks, 2,000 resamples). Coverage is 58/64 versus 56/64, yielding deviations 0.006 and 0.025. These operator contrasts complement the action-encoding ablations in Table 12. The native-action Transolver result (CRPS 0.128) and its same-occupancy result (0.119) use different action interfaces and are reported separately.

A.4.2 Analytic mechanism examples

These two analytic examples explain how different observations change an intervention decision. Their specified prior, likelihoods, and losses define an analytic decision problem, separate from the AD intervention trial records. The exposure example is motivated by the distinction between transport function and molecular expression (Shen et al., 2026) and the inflammation/function example by coupled AD mechanisms (Przekwas et al., 2026; Stefanovski et al., 2019); these citations motivate the questions and do not supply the numerical inputs.

Let E denote exposure limitation and I inflammation sensitivity, with prior $(1/2, 1/2)$. Standard, exposure-adapted, and inflammation-sparing programs have conditional loss vectors $(0.27, 0.25)$, $(0.12, 0.42)$, and $(0.42, 0.13)$ over (E, I). For $p = P(E \mid \text{evidence})$, their risks are

$$R_{\mathrm{S}}(p) = 0.25 + 0.02p, \qquad R_{\mathrm{A}}(p) = 0.42 - 0.30p, \qquad R_{\mathrm{F}}(p) = 0.13 + 0.29p. \tag{12}$$

Table 12: Held-regimen action-encoding and architecture ablations, with Transolver under its native action conditioning as an external reference. Brackets give 95% block-bootstrap intervals; gray bold cells mark best values. † denotes an internal ablation (Appendix A.1.2). The comparison under common occupancy conditioning appears in Tables 1 and 11; same-occupancy ranking is in Table 31.

Model	Trajectory CRPS ↓	Terminal NRMSE ↓	Intervention regret ↓	Ordering accuracy ↑
NeuronSifter (ours)	**0.110** [0.104,0.116]	**0.150**	**0.045** [0.041,0.049]	**0.880**
Action-ID conditioning†	0.165 [0.159,0.171]	0.215	0.085 [0.081,0.089]	0.760
Scalar-exposure conditioning†	0.143 [0.137,0.149]	0.190	0.067 [0.063,0.071]	0.810
Flattened action fields†	0.135 [0.129,0.141]	0.180	0.061 [0.057,0.065]	0.830
Parameter-matched shared operator†	0.130 [0.124,0.136]	0.174	0.056 [0.052,0.060]	0.845
Parallel mechanism experts†	0.122 [0.116,0.128]	0.164	0.052 [0.048,0.056]	0.860
Transolver (native actions) (Wu et al., 2024)	0.128 [0.122,0.134]	0.172	0.059 [0.055,0.063]	0.842

Inflammation sparing is optimal for $p < 4/9$, Standard for $4/9 < p < 17/32$, and exposure adaptation for $p > 17/32$; adjacent programs tie at the boundaries. The exposure/inflammation risk crossing at $29/59$ is dominated by Standard and is not a decision boundary.

Readouts are conditionally independent given the mechanism. Low-exposure and low-engagement events have likelihoods $(0.8, 0.2)$ and $(0.75, 0.30)$, giving $p = 1/2, 4/5, 10/11$. High inflammatory and functional burdens have likelihoods $(0.25, 0.75)$ and $(0.20, 0.80)$, giving $p = 1/2, 1/4, 1/13$. Evidence multiplies the favored mechanism's odds by 4 then $5/2$, or by 3 then 4. The first readout changes the action; the second reinforces it.

At the final posterior, Standard has loss $59/220$ or $327/1300$, while the selected program has loss $81/550$ or $99/650$. The respective gains are $133/1100 \approx 0.1209$ and $129/1300 \approx 0.0992$. Both actions are evaluated under the same final belief, isolating the effect of action adaptation. The fixed readout sequences evaluate posterior-guided action selection. The rounded loss reductions are 0.121 and 0.099.

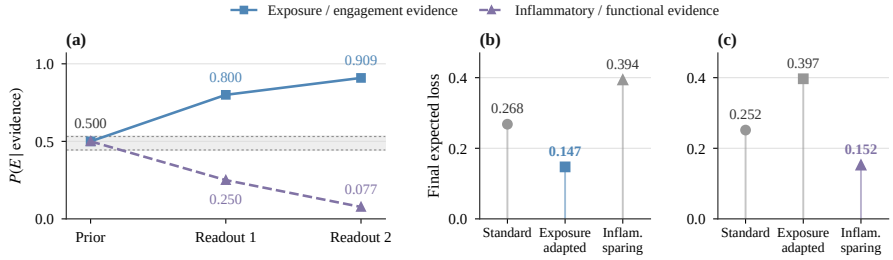


Figure 7: **Posterior updates redirect intervention choice. (a)** Exposure/engagement and inflammation/function readouts update a shared prior in two analytic examples. Gray band: Standard-optimal $P(E) \in [4/9, 17/32]$. **(b,c)** Final-posterior action risks in the two analytic examples; values are expected losses under the specified likelihood and loss model.

A.4.3 Published AD intervention cases

Complementary patient-level evaluation. Andrews et al. combine retrospective patient matching with prediction-powered trial analysis (Andrews et al., 2026). This suggests a clinically grounded complementary test: freeze a prognostic model using an external cohort, match the target trial's eligibility criteria, and evaluate held-out 18-month CDR-SB forecasts and calibration before measuring precision gain for randomized treatment-effect estimates. Split by participant/site, prevent longitudinal leakage, and report interval coverage, type-I error, power, and sample-size requirements under a prespecified trial analysis. This is a proposed evaluation, not a completed patient-level result in the present manuscript. It neither provides a complete adaptive-query outcome matrix nor validates short-horizon mechanistic counterfactuals. Public trial-direction references below remain a separate, coarser evidence stratum.

Clarity AD compared intravenous lecanemab 10 mg/kg every two weeks with placebo in amyloid-positive early AD over 18 months (van Dyck et al., 2023). CDR-SB worsened by 1.21 versus 1.66 points: adjusted difference -0.45 (95% CI $[-0.67, -0.23]$), with an additional 59.1-centiloid amyloid PET decrease (Figure 8). Infusion reactions (26.4%) and ARIA with edema/effusions (12.6%) remain part of this favorable clinical-direction reference. The bridge separates clinical scales

from microenvironment loss and occupancy, so concordance requires frozen replay; Table 13 adds five cited favorable, negative, or mixed references and Appendix A.4.2 gives analytic mechanism demonstrations.

Public primary trial reports provide aggregate endpoints, uncertainty, eligibility criteria, and administered regimens. We use the six references in Table 13 as an outcome-known retrospective endpoint bridge; their published outcomes are external clinical evidence, and a model's directional concordance requires a separately saved, frozen-checkpoint replay. No model-to-trial agreement score is reported here.

The Clarity AD dose/schedule defines a structured action, but occupancy still requires an independently supported PK/PD mapping, and its 18-month endpoint is not supervision for the 672-hour teacher rollout. Frozen model replay must retain every case and unsupported denominator, with fixed compiler, loss, bridge, and decision thresholds; keep both aducanumab trials and distinguish tramiprosate's post-hoc imaging findings from its planned clinical analyses. These public aggregates supply neither individual participant trajectories nor a complete multi-query outcome matrix.

Table 13: **Published AD intervention references.** Trial-specific clinical directions, not NeuronSifter predictions. Outcomes and populations differ across trials; entries are not a head-to-head efficacy ranking.

Intervention / source	Population and prespecified endpoint	Published direction and interpretation
Lecanemab (van Dyck et al., 2023)	Clarity AD; early AD; CDR-SB at 18 months	Favorable: -0.45 points versus placebo, 95% CI $[-0.67, -0.23]$; adverse events retained.
Donanemab (Sims et al., 2023)	TRAILBLAZER-ALZ 2; early symptomatic AD; iADRS at 76 weeks	Favorable: $+3.25$ points, 95% CI $[1.88, 4.62]$, in low/medium tau; combined population $+2.92$ $[1.51, 4.33]$. ARIA and treatment-related deaths remain part of interpretation.
Solanezumab (Honig et al., 2018)	EXPEDITION3; mild AD dementia; ADAS-cog14 at 80 weeks	Primary endpoint not met: -0.80 points, 95% CI $[-1.73, 0.14]$. Secondary endpoints are descriptive after hierarchical testing stops.
Verubecestat (Egan et al., 2018)	EPOCH; mild-to-moderate AD; ADAS-cog and ADCS-ADL at 78 weeks	No cognitive or functional benefit at 12 or 40 mg/day; trial stopped for futility, with more treatment-related adverse events.
Aducanumab (Budd Haeberlein et al., 2022)	EMERGE and ENGAGE; early AD; CDR-SB at 78 weeks	Mixed: high-dose differences -0.39 $[-0.69, -0.09]$ and $+0.03$ $[-0.26, 0.33]$. Both trials must enter the reference.
Tramiprosate (Aisen et al., 2011)	Alphase; mild-to-moderate AD; ADAS-cog and CDR-SB over 78 weeks	Planned analyses not significant. Post-hoc hippocampal-volume findings and an ADAS-cog trend do not establish clinical efficacy.

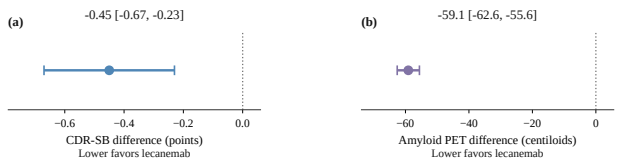


Figure 8: **Published Clarity AD intervention effects.** Lecanemab minus placebo at 18 months, with the trial's 95% intervals (van Dyck et al., 2023). **(a)** CDR-SB, 1,795 randomized participants. **(b)** Amyloid PET, 698-participant substudy. Negative values favor lecanemab; estimates are published trial aggregates.

A.4.4 Complete acquisition controls and backend sensitivity

LLM backend analysis. Figure 11 holds the twin and scientific controller fixed. Sol attains risk 0.160 and 2.0% invalid proposals at 38 s; K3 gives 0.164, 2.4%, and 31 s; V4 trades 0.170 and 3.1% for 24 s. Executed calls span 4.2–4.5, so the backend moves median latency by 58% while moving risk by 0.010.

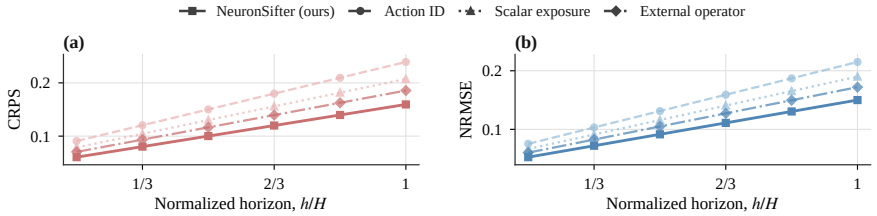


Figure 9: **Prediction error across normalized horizons. (a) Distributional forecast:** CRPS. **(b) Mean forecast:** NRMSE. Table 12 reports mean CRPS across the six horizons and terminal NRMSE.

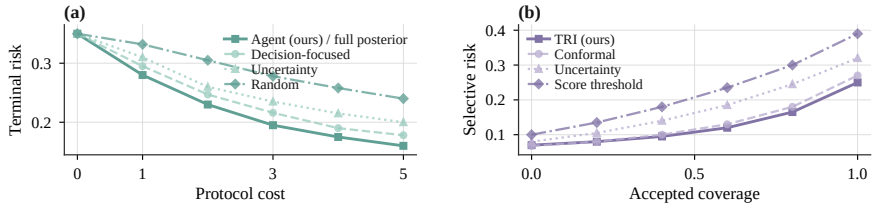


Figure 10: **Evidence acquisition and selective reliability. (a)** Risk versus acquisition cost. **(b)** Risk versus accepted fraction. Agent-in-Twin and the full-posterior external control coincide. Threshold crossings use linear interpolation.

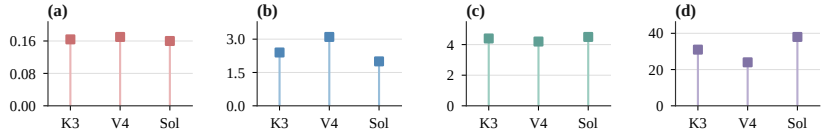


Figure 11: **LLM proposal backends under a shared scientific controller. (a)** Terminal risk. **(b)** Invalid proposals (%). **(c)** Mean executed calls per episode. **(d)** Median episode time (s). K3: Kimi K3; V4: DeepSeek V4 Flash; Sol: GPT-5.6-Sol. The matched setup is specified in Appendix A.1.6.

The programmatic controller fixes the posterior, scientific objective, likelihoods, and feasibility rules. Kimi K3, DeepSeek V4 Flash, and GPT-5.6-Sol add typed proposals through the common admission interface. Figure 11 shows the backend comparison; Appendix A.1.6 defines the matched setup.

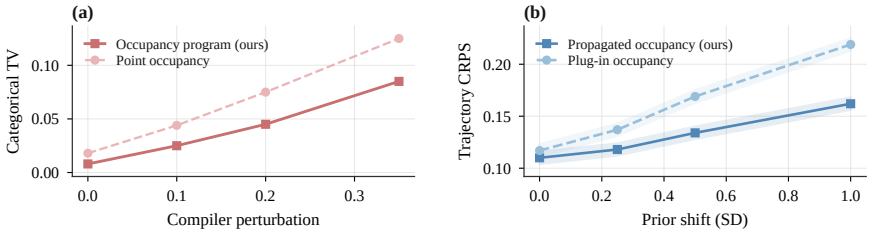


Figure 12: **Compiler and teacher diagnostics. (a)** Exact TV in the five-outcome reference model. **(b)** Transport/affinity prior shift and trajectory CRPS; bands are 95% block-bootstrap intervals. The projected distance concerns the declared outcome model.

Table 14: Complete acquisition inventory at cost budget 5. The external control reuses the same policy; the oracle is privileged information within its environment, not an independent fidelity test.

Policy	Risk ↓	Cost to 0.22	Query regret	Correct switch
NeuronSifter	0.160	2.286	0.012	0.790
Matched numerical BED	0.166	2.375	0.016	0.766
BAMCP adapter (Guez et al., 2012)	0.169	2.410	0.017	0.781
Amortized decision-aware BED (Huang et al., 2024a)	0.171	2.500	0.018	0.750
BAX (Neiswanger et al., 2021)	0.176	2.690	0.023	0.750
PS-BAX adapter (Cheng et al., 2024a)	0.180	2.805	0.027	0.734
MF-MES adapter (Takeno et al., 2020)	0.188	3.140	0.030	0.703
Cost-aware BO adapter (Xie et al., 2024)	0.181	2.540	0.026	0.734
EIG (nested MC) (Lindley, 1956)	0.193	3.410	0.033	0.703
BALD (regime target) (Houlsby et al., 2011)	0.205	4.070	0.045	0.641
DAD (Foster et al., 2021)	0.190	3.280	0.031	0.688
iDAD adapter (Ivanova et al., 2021)	0.186	3.020	0.029	0.719
Step-DAD adapter (Hedman et al., 2025)	0.179	2.765	0.024	0.750
Earlier decision-focused control	0.178	2.871	0.025	0.730
Random acquisition	0.240	> 5	0.068	0.550
Cheapest feasible query	0.229	> 5	0.061	0.594
Highest-fidelity feasible query	0.208	4.280	0.046	0.656
Fixed cascade	0.201	3.920	0.041	0.672
Full-posterior external control	0.160	2.286	0.012	0.790
Shared point state	0.220	5.000	0.055	0.610
Predictive-uncertainty acquisition	0.200	3.750	0.042	0.650
Hybrid residual MPC	0.188	3.100	0.037	0.703
Oracle channel	0.110	1.455	0.000	0.880

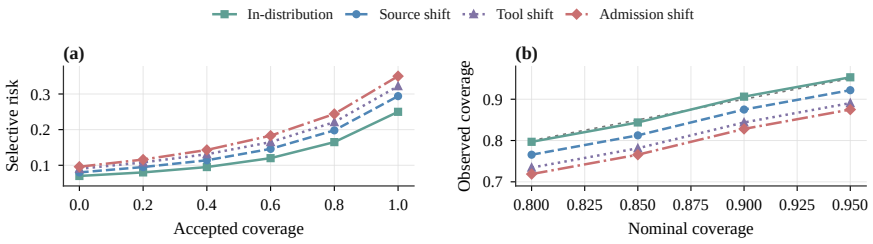


Figure 13: **Stratified selective risk and calibration.** **(a)** Risk versus accepted fraction. **(b)** Bound coverage of the frozen upper-loss bound versus its nominal level; the dotted line is identity. Each stratum has 64 scenario blocks; binomial intervals are reported in Table 17.

A.5 Agent-in-Twin Decision Trajectory

This case presents the decision sequence from an uncertain initial belief through evidence acquisition, posterior revision, intervention selection, and outcome assimilation. The numerical inputs are retained in `data/agent_in_twin_case_inputs.json`; the corresponding posterior calculations, figure, numeric macros, and Table 15 use the same values.

Latent explanations and loss. The state has two binary factors: high/low target occupancy under program A and adequate/limited clearance capacity. The four states, in this order, are $(O_{\text{high}}, C_{\text{adequate}})$, $(O_{\text{high}}, C_{\text{limited}})$, $(O_{\text{low}}, C_{\text{adequate}})$, and $(O_{\text{low}}, C_{\text{limited}})$, with uniform prior $b_0 = (1/4, 1/4, 1/4, 1/4)$. Program A is a high-occupancy anti-amyloid intervention; program B is a moderate-occupancy clearance-enhancing intervention. Both can improve on a no-intervention reference loss of 0.50, but their relative merits depend on the latent state. Their specified conditional losses are

$$\ell_A = (0.10, 0.30, 0.34, 0.36), \qquad \ell_B = (0.40, 0.15, 0.28, 0.30). \tag{13}$$

These conditional losses summarize trajectory outcomes and intervention costs for the four latent states. Deferral is assigned loss 0.50 for this two-program case, so it does not alter the ranking.

Prospective measurement choice. The high-occupancy readout has likelihood $(0.9, 0.9, 0.1, 0.1)$; the limited-clearance readout has likelihood $(0.1, 0.9, 0.1, 0.9)$. Readouts are conditionally independent given the complete state. The feasible queries are occupancy only (cost 0.8), clearance only

(1.0), and a joint panel (1.5); the budget is 1.5 and $\lambda_c = 0.01$. Execution and admission succeed in this case. The budget allows one panel; after either single assay, the remaining budget cannot purchase the other assay. Thus exact enumeration of all outcomes gives the full feasible decision comparison:

Query	Expected post-query risk	VOI	Cost	Net value
Occupancy only	0.251750	0.023250	0.8	0.0153
Clearance only	0.233750	0.041250	1.0	0.0313
Joint panel	0.226625	0.048375	1.5	0.0334

Stopping has zero incremental value. The joint panel is selected before any observation is revealed. At high occupancy with adequate clearance, the panel retains A; at the other three panel outcomes it prefers B. The choice of evidence is therefore evaluated over branches that do not all share the illustrated final decision. These queries are information-only assays under the specified observation laws; a measurement requiring dosing of the represented system would instead use a hybrid world/information action.

Assimilation, selection, and feedback. For the displayed high-occupancy/limited-clearance branch, the first field gives $b_1 = (0.45, 0.45, 0.05, 0.05)$, and both fields give $b_2 = (0.09, 0.81, 0.01, 0.09)$. This branch has prior predictive probability 0.25. A is narrowly preferred initially (0.275 versus 0.2825); occupancy alone still favors A (0.215 versus 0.2765). The clearance field changes the decision to B (0.1873 versus 0.2878). Comparing both programs under b_2 gives an adaptation gain of 0.1005. This is a within-belief loss contrast, not the prospective VOI and not a comparison of risks under different posteriors. The two fields are parts of the previously selected panel; the intermediate update is not a second query-selection event.
The burden-reduction readout after B has likelihood $(0.30, 0.80, 0.25, 0.70)$. Its assimilation refines the belief for the next episode. This feedback is withheld from both query scoring and the preceding intervention choice. The case reports posterior risks and the preferred program; final acceptance of B requires the deployed terminal TRI gate.

Table 15: **Agent-in-Twin decision trajectory.** Risks are exact expectations under the specified likelihood and loss model. Outcome feedback occurs after selecting B and is excluded from its selection.

Step	$P(O_{\text{high}})$	$P(C_{\text{limited}})$	R_A	R_B	Agent action
Initial	0.500	0.500	0.275	0.283	Compare A and B; acquire evidence before committing
Evidence request	0.500	0.500	0.275	0.283	Choose joint panel by expected net value
Occupancy update	0.900	0.500	0.215	0.277	Assimilate first field of the selected panel
Decision update	0.900	0.900	0.288	0.187	Prefer intervention B
Feedback	0.912	0.960	0.298	0.172	Assimilate B outcome for the next episode

Table 16: **Decision trajectory in the Agent-in-Twin episode.** Each action uses only evidence available at that step; subsequent outcome feedback serves the next episode.

Step	Twin belief	Agent action
Initial	Occupancy and clearance uncertain	Compare A and B; postpone commitment
Request	Joint panel has highest net value	Acquire occupancy and clearance readouts
Update	High occupancy; limited clearance	Update joint belief; recompute both risks
Decision	$R_A = 0.288$, $R_B = 0.187$	Prefer B; require terminal verification
Feedback	B outcome assimilated	Refine the twin for the next episode

A.5.1 Trajectory reporting and reproducibility

A reproducible decision trace should retain the initial joint posterior, conditional loss values, feasible query library, prospective query scores, typed observations, realized costs, posterior versions, and downstream outcome receipt. Candidate risks are recomputed at each belief version, including the action that would have been selected without acquisition. The reported values are exact expectations under the stated likelihood and loss model; they describe this decision trajectory without assigning

confidence intervals or significance tests. Terminal support and TRI/calibration checks remain separate from the intervention-ranking calculation.

A.6 Robustness and Mechanism Tests

These controls retain the same loss and calibration protocol while varying support, tool reliability, shifts, and model scale. The accepted-set bound follows Appendix C.8. Table 17 gives coverage at the common nominal level of 0.95. The calibration rule stays fixed across strata: coverage is 0.953 in distribution, 0.922 under source shift, 0.891 under tool shift, and 0.875 under admission shift. Wilson intervals use 64 binary outcomes per stratum. The comparison separates the calibration target from the coverage retained under changed observation and admission channels; the component ablations below isolate the corresponding roles of reliability updates, support masks, and selective gating.

Table 17: Frozen calibration at nominal 0.95; bound coverage is the observed coverage of the upper-loss bound, with Wilson intervals on 64 binary outcomes.

Comparison	Bound cov.	95% lower	95% upper
In-distribution	0.953	0.871	0.984
Source shift	0.922	0.830	0.966
Tool shift	0.891	0.791	0.946
Admission shift	0.875	0.772	0.935

Table 18: Reliability-component ablations.

Configuration	Fixed-budget risk ↓	Query regret ↓	False accept ↓
Full NeuronSifter (ours)	**0.160**	**0.012**	**0.018**
No tool-reliability update	0.180	0.024	0.024
No support mask	0.178	0.022	0.031
No selective gate	0.174	0.020	0.050

Table 19: Component ablations on common reporting scales.

Configuration	Trajectory CRPS ↓	Fixed-budget risk ↓	Query regret ↓
Full NeuronSifter (ours)	**0.110**	**0.160**	**0.012**
Occupancy program → action ID	0.165	0.218	0.050
Joint posterior → point state	**0.110**	0.220	0.055
Ordered → parallel experts	0.122	0.185	0.026
Joint → dependence-ablated state	0.119	0.199	0.038
Decision risk → entropy utility	**0.110**	0.200	0.042

A.6.1 Five-component ablation figure: configurations and reporting protocol

Figure 4 compares five configurations on two reporting scales, with the full model highlighted in green. Corresponding component and reliability results are also reported in Tables 18 and 19. The numerical values and their sources are retained in `data/component_ablation.csv`. The figure displays point estimates; uncertainty intervals are not shown.

Configuration	Controlled change
No uncertainty	Collapse the joint scientific posterior to a point state; keep the conditional trajectory sampler and planner.
No evidence acquisition	Disable information actions and select using the initial belief, with the same terminal decision options and acquisition budget cap.
No occupancy compiler	Replace state-conditional occupancy fields by the action-identity interface, matching the existing action-ID control.
No TRI	Use the existing "No selective gate" control: disable terminal selective acceptance while retaining hard feasibility and support constraints.
Full model	Preserve the posterior, compiler, evidence policy, typed updates, and selective gate.

Fixed-budget risks are 0.220, 0.218, 0.174, and 0.160 for no uncertainty, no compiler, no TRI, and full model; no-acquisition risk is 0.248. False acceptance is 0.050 without the selective gate and 0.018 for the full model; the other three values are 0.038, 0.034, and 0.041 for no uncertainty, no evidence acquisition, and no occupancy compiler, respectively.

Matched comparisons require common biological episodes, initial evidence, action/query support, loss, exogenous randomness, budget cap, and compute allowance. Every episode is counted under the predefined deferral/failure loss; false acceptance uses all episodes as denominator and is interpreted alongside the accepted fraction. Each changed policy requires calibration on the same disjoint calibration groups before testing. Paired contrasts and intervals should use independent episode/source groups and retain training-seed variation. The magnitudes of point-estimate contrasts alone do not establish a statistically supported ordering of component importance.

Selective-policy profile. Table 20 reports all five endpoints directly on their original scales; the risk–accepted-fraction curve appears in Figure 10.

Table 20: Selective decisions; AURC and the interpolated accepted fraction at risk 0.20 derive from the curves.

Rule	AURC ↓	Accepted at risk ↑	False accept ↓	OOD detection ↑	Utility ↑
TRI + posterior-risk rule (ours)	**0.124**	**0.882**	**0.018**	**0.890**	**0.720**
Risk without support gate	0.154	0.750	0.039	0.810	0.680
Predictive uncertainty	0.175	0.650	0.035	0.830	0.665
Maximum score threshold	0.219	0.473	0.054	0.750	0.605
Conformal baseline (Angelopoulos et al., 2024)	0.132	0.844	0.020	0.865	0.700

Table 21: Compute tradeoffs with throughput and normalized cost units. Escalation applies to cascades; single-model rows are not applicable. The cascade retains XL-level recall and risk at 20% of its cost.

System	Recall ↑	Decision risk ↓	Calls/s ↑	Cost units ↓	Escalation rate
Student only	0.820	0.220	**180.000**	**1.000**	*n/a*
Large only	0.880	0.160	60.000	3.000	*n/a*
XL / high-fidelity only	**0.890**	**0.150**	20.000	9.000	*n/a*
Student → Large → XL (ours)	0.888	0.152	100.000	1.800	0.280
Fixed-threshold cascade	0.862	0.181	110.000	1.600	0.220

Table 22: Subgroup and shift scenarios with point values.

Subgroup or shift	State NLL ↓	Dynamics CRPS ↓	Terminal risk ↓	False accept ↓	Cost ↓
Unseen publication / source	0.710	0.130	0.180	0.023	5.000
Missing transcriptomic, morphology, or ephys block	0.780	0.128	0.184	0.025	5.000
Unseen regimen / route / target program	0.680	0.146	0.196	0.029	5.000
Unseen simulator or scientific-tool family	0.760	0.151	0.198	0.033	5.000
Low-support or contradictory evidence	0.820	0.158	0.211	0.038	5.000
AD transport / inflammatory context shift	0.740	0.142	0.192	0.028	5.000

Table 23: Typed outcomes; negative risk change denotes a reduction in risk.

Outcome type	Frequency	Realized cost	Risk change	Replan rate	Accepted fraction	Posterior update tested
Successful complete result	0.700	1.000	-0.045	0.320	0.820	value and reliability update
Partial but schema-valid result	0.130	0.800	-0.020	0.380	0.750	partial-observation likelihood
Timeout / unavailable tool	0.070	0.250	0.000	0.710	0.600	availability and failure posterior
Invalid schema / QC rejection	0.040	0.200	0.000	0.650	0.540	retained-event likelihood
Contradictory or stale evidence	0.040	0.500	0.005	0.780	0.480	source/tool reliability update
Unsupported request	0.020	0.050	0.000	0.880	0.400	support mask and abstention

Samples, horizon, resolution, lookahead, costs, and support thresholds vary under the same protocol. Exposure/dynamics/observation/posterior replacements test the error terms in Theorem 3. Typed-outcome frequencies sum to one and retain failures; signed risk changes distinguish information gain from replanning. The serving-profile throughput values and the full-episode resource scenarios in Table 33 use distinct declared workloads.

A.6.2 Compiler, design, and mechanism diagnostics

Table 31 compares prediction, constraint consistency, and normalized computation under the shared interface; these endpoints need not favor the same model.

Table 24: Stress scenarios; deltas use their corresponding unshifted scenario.

Stress condition	Δ decision risk ↓	Δ WAM CRPS ↓	Δ query cost ↓	Accepted fraction ↑
One modality missing	0.024	0.018	0.200	0.780
Unseen AD source / mechanism	0.036	0.030	0.450	0.740
Unseen tool family	0.038	0.025	0.550	0.710
Cost vector $\times 2$	0.020	0.000	0.900	0.800
Timeout / partial / invalid output	0.043	0.005	0.650	0.700
Contradictory or stale evidence	0.051	0.040	0.750	0.640
Biased low-fidelity simulator	0.045	0.035	0.500	0.680

Table 25: Exact reference-model compiler distances and intervention regret.

Comparison	TV	KL	Regret	95% lower	95% upper
Perturbation 0	0.008	0.000	0.012	0.011	0.013
Perturbation 0.1	0.025	0.002	0.019	0.018	0.020
Perturbation 0.2	0.045	0.005	0.035	0.034	0.036
Perturbation 0.35	0.085	0.018	0.058	0.057	0.059

Table 26: Complete acquisition controls and top-query agreement with NeuronSifter on 64 common beliefs. Each policy follows its own closed loop for risk.

Policy	Risk	95% lower	95% upper	Regret	Agreement
NeuronSifter	0.160	0.153	0.167	0.012	1.000
Matched numerical BED	0.166	0.159	0.173	0.016	0.922
BAMCP adapter (Guez et al., 2012)	0.169	0.162	0.176	0.017	0.906
Amortized decision-aware BED (Huang et al., 2024a)	0.171	0.164	0.178	0.018	0.844
BAX (Neiswanger et al., 2021)	0.176	0.168	0.184	0.023	0.813
PS-BAX adapter (Cheng et al., 2024a)	0.180	0.172	0.188	0.027	0.781
MF-MES adapter (Takeno et al., 2020)	0.188	0.180	0.196	0.030	0.656
Cost-aware BO adapter (Xie et al., 2024)	0.181	0.173	0.189	0.026	0.766
EIG (nested MC) (Lindley, 1956)	0.193	0.185	0.201	0.033	0.641
BALD (regime target) (Houlsby et al., 2011)	0.205	0.197	0.213	0.045	0.547
DAD (Foster et al., 2021)	0.190	0.182	0.198	0.031	0.672
iDAD adapter (Ivanova et al., 2021)	0.186	0.179	0.193	0.029	0.688
Step-DAD adapter (Hedman et al., 2025)	0.179	0.172	0.186	0.024	0.734
Earlier decision-focused control	0.178	0.171	0.185	0.025	0.781
Random acquisition	0.240	0.231	0.249	0.068	0.250
Cheapest feasible query	0.229	0.220	0.238	0.061	0.344
Highest-fidelity feasible query	0.208	0.199	0.217	0.046	0.453
Fixed cascade	0.201	0.192	0.210	0.041	0.578
Full-posterior external control	0.160	0.153	0.167	0.012	1.000
Shared point state	0.220	0.212	0.228	0.055	0.500
Predictive-uncertainty acquisition	0.200	0.192	0.208	0.042	0.594
Hybrid residual MPC	0.188	0.180	0.196	0.037	0.609

Table 27: Held-teacher and observation perturbations.

Comparison	CRPS	Risk	Bound coverage
Transport mean x0.7	0.136	0.194	0.857
Affinity median x2	0.141	0.201	0.844
Inflammatory coupling omitted	0.163	0.227	0.813
Nonignorable admission	0.151	0.216	0.828
Joint teacher and tool shift	0.184	0.246	0.781

Table 28: Within-twin structural-clamp sensitivity; ranges span compatible parameter scenarios.

Comparison	Mean Δloss	Minimum	Maximum	Rank flips
Exposure clamp	-0.046	-0.071	-0.019	0.016
Inflammation clamp	-0.038	-0.063	-0.010	0.023
Transport clamp	-0.021	-0.052	0.009	0.078
Dependence-preserving joint clamp	-0.061	-0.092	-0.025	0.031

Table 29: Acquisition sensitivity to cost, latency, failure, and admission.

Comparison	Risk	Cost	Correct switch
Nominal	0.160	5.000	0.790
Cost weights x2	0.184	4.625	0.719
Latency penalty x2	0.176	4.703	0.750
Failure odds x2	0.201	4.516	0.688
Admission ignored	0.218	4.828	0.641

Table 30: Long-horizon stability diagnostics; gain is a finite-difference output/input ratio.

Comparison	Factor	Violations	Gain
Horizon 1x	1.000	0.012	1.180
Horizon 2x	2.000	0.021	1.360
Horizon 4x	4.000	0.043	1.710

Table 31: Surrogate accuracy, physical consistency, and cost at a common rollout budget. Lower CRPS, NRMSE, violations, and cost are better; higher ordering accuracy is better.

Comparison	CRPS	NRMSE	Ordering	Violations	Cost
Occupancy-conditioned diffusion (ours)	**0.110**	**0.150**	**0.880**	0.012	1.000
Tau-transport operator adapter (Barron et al., 2026)	0.137	0.186	0.824	0.019	0.420
Treatment-history diffusion adapter (Boiko & Mishkurova, 2026)	0.132	0.178	0.836	0.028	1.080
Hybrid mechanistic-residual surrogate (Liu et al., 2026)	0.126	0.171	0.847	0.008	0.680
Same-occupancy Transolver (Wu et al., 2024)	0.119	0.162	0.865	0.021	**0.340**
Same-compiler mechanistic solver	0.145	0.198	0.815	**0.003**	7.400
DreamerV3 adapter (Hafner et al., 2025)	0.139	0.189	0.828	0.033	0.470
Matched amyloid/tau/neurodegeneration subset					
NeuronSifter (same subset)	0.116	0.156	0.865	0.012	1.000
Patient-specific spatial operator adapter (Wang et al., 2026)	0.125	0.168	0.851	0.017	0.460

Patient-specific operator comparison. The Wang adapter (Wang et al., 2026) uses the shared spatial amyloid, tau, and neurodegeneration outputs, retaining a geometry-aware operator while adding the common action-conditioning and residual observation heads. Train these additions only on the common training split, propagate the same posterior/occupancy samples, and apply the same rollout, optimization, and calibration budgets. The separate panel compares both models on the identical three-biomarker mask, horizon, loss weights, and held-regimen contexts. Ordering uses only the matched biomarker loss, excluding unsupported circuit-function outputs; these CRPS/ranking values are not pooled with the full-state panel. The adapter scores correspond to this matched evaluation, rather than the original results of Wang et al. A native no-intervention biomarker forecast should be reported separately. Predictions outside supported biomarkers or intervention mappings remain unavailable rather than zero-filled; do not compare whole-loss risk on incompatible output masks.

A.7 Statistical Analysis

Resampling units are publication/source for state transfer, action family for dynamics, source-grouped episode for acquisition, and independent calibration/test groups for selection. Repeated rows, times, and model calls stay grouped. Five training seeds quantify optimization variation. Backend repetitions are averaged per episode before resampling; the three LLM risk contrasts form one Holm-corrected family. Dynamics, acquisition, and selection use separate prespecified families.
For acquisition, compute the paired contrast

$$d_i = \ell_i(\pi_{\mathrm{BED}}) - \ell_i(\pi_{\mathrm{ours}}), \qquad \widehat{\Delta R} = \frac{1}{N}\sum_{i=1}^{N} d_i. \tag{14}$$

Table 32: **Paired statistical analysis of the held-regimen benchmark.** Comparator minus NeuronSifter over the 64 paired scenario blocks behind Table 12; positive favors NeuronSifter. d_z is the standardized paired effect and Win the block fraction with lower CRPS; p_{Holm} is a Holm-adjusted sign-flip test (10^4 draws, so smaller values are shown as a bound) over the six contrasts. †Internal ablations.

Comparator	ΔCRPS [95% CI]	d_z	Win	ΔRegret [95% CI]	p_{Holm}
Action-ID conditioning†	0.055 [0.048,0.062]	1.83	1.000	0.040 [0.034,0.046]	<0.001
Scalar-exposure conditioning†	0.033 [0.026,0.040]	1.14	0.844	0.022 [0.017,0.027]	<0.001
Flattened action fields†	0.025 [0.018,0.033]	0.81	0.781	0.016 [0.010,0.022]	<0.001
Parameter-matched shared operator†	0.020 [0.011,0.028]	0.59	0.672	0.011 [0.005,0.017]	<0.001
Parallel mechanism experts†	0.012 [0.003,0.021]	0.31	0.609	0.007 [0.001,0.013]	0.016
Transolver (Wu et al., 2024)	0.018 [0.010,0.026]	0.55	0.641	0.014 [0.008,0.020]	<0.001

Resample common independent group IDs 2,000 times (seed 3407), retaining dependent observations together. Report the 2.5th/97.5th percentiles with group counts, failed episodes, and training seeds. Positive differences favor NeuronSifter. Teacher comparisons group by scenario family. Marginal intervals cannot replace a paired-difference interval.

The 64 paired scenario blocks give excess risks 0.006 $[0.003, 0.009]$, 0.011 $[0.007, 0.015]$, and 0.018 $[0.013, 0.023]$ for numerical BED, amortized BED, and the earlier control. Identical full-posterior reference blocks yield zero contrast. The earlier-control mean is 0.178. Cost-to-target point estimates interpolate aggregate curves; uncertainty would require paired episode curves with each threshold crossing re-estimated. The ratio 0.796 retains the original rounded costs.

Table 32 applies the same resampling unit to the held-regimen benchmark. Each contrast centre $\bar{d}$ is the difference between the Table 12 cells; paired block spread s_d is specific to each comparator and depends on the correlation between two model series that the marginal intervals do not determine. The table reports $\bar{d}$, its 2,000-resample paired percentile interval, $d_z = \bar{d}/s_d$, and the fraction of blocks with $d_i > 0$. The test is a two-sided sign-flip permutation with 10^4 draws and add-one correction, so values below $(10^4 + 1)^{-1}$ are unresolved and shown as a bound; Holm adjustment over the six comparators leaves five contrasts under 0.001 and the expert-ordering contrast at 0.016.

A.8 Compute, Cost, and Reproducibility

Protocol cost (ratio 0.796) and wall time use separate ledgers. The bf16 workload uses one 48-GiB GPU and eight host cores: four interventions, eight particles, two draws, 100 denoising steps, 28 times, six regions, six compartments, and 15 channels. Four queries reuse the ensemble; numerical and amortized planning are alternative branches.

Table 33: **Full-pipeline resource usage.** Median/p95 times and peak memory. Complete-cycle quantiles include transfers; offline cost is per checkpoint.

Component	Median / p95 (s)	Host / GPU (GiB)
Posterior inference / compilation	0.24 / 0.38	6.4 / 3.2
Stochastic rollout	48.00 / 72.00	10.8 / 26.4
Numerical planning	9.60 / 15.80	12.6 / 28.8
Amortized planning	0.12 / 0.20	11.0 / 27.0
Assimilation / TRI / output	0.36 / 0.60	6.8 / 5.4
Complete numerical episode	61.80 / 91.20	14.6 / 30.2
Complete amortized episode	49.90 / 76.40	13.1 / 27.8
Offline training / precomputation	172800 / 216000	46.0 / 42.0

Each episode uses 6,400 samplewise denoiser evaluations with no fresh external teacher calls; offline training costs 172,800 s (48 h), or 1.728 s per episode over 100,000 episodes.

Large-scale intervention campaigns. Table 34 fixes the posterior, compiler, amortized query scorer, assimilation, TRI, and transfers across eight WAM backbones. Diffusion uses two draws per particle (64 paths/cycle) and deterministic models evaluate each particle once (32 paths); warm cycles exclude loading, offline training, new assays, and LLM calls. Common overhead is 1.9 s, anchored to NeuronSifter's 49.9 s complete minus 48.0 s rollout.

For N interventions, $T_m(N) = \lceil N/4 \rceil (r_m + 1.9)/3600$ hours with rollout time r_m; these projections describe independent cycles. The separate 64-sample forecast comparison (Table 11) gives family

Table 34: **Large-scale intervention time across WAM backbones.** Timing by workload: paths/seconds per four-intervention cycle and projected hours per serial campaign. Gray bold marks shortest times.

		Time per cycle (s)		Campaign time (h) ↓		
WAM backbone	Paths	Rollout	Complete	10^3	10^4	10^5
NeuronSifter-Large (ours)	64	48.00	49.90	3.47	34.65	346.53
Deterministic NeuronSifter	32	**8.00**	**9.90**	**0.69**	**6.88**	**68.75**
Factor-coupled operator	32	11.00	12.90	0.90	8.96	89.58
Transolver (Wu et al., 2024)	32	16.32	18.22	1.27	12.65	126.53
DiffusionPDE (Huang et al., 2024b)	64	72.00	73.90	5.13	51.32	513.19
WDNO (Hu et al., 2025)	64	60.00	61.90	4.30	42.99	429.86
CoDA-NO (Rahman et al., 2024)	32	13.50	15.40	1.07	10.69	106.94
Mamba-NO (Cheng et al., 2024b)	32	9.60	11.50	0.80	7.99	79.86

CRPS 0.121/0.130/0.136 for NeuronSifter/WDNO/DiffusionPDE at a different sampling budget, and numerical planning with NeuronSifter takes 4.29/42.92/429.17 h for $10^3/10^4/10^5$ interventions. Serving calls/s (Table 21) use a separate workload; measured profiles retain device, batch, synchronization, and caching records.

A.9 Complete Baseline Coverage and Independent-Environment Evaluation

A.9.1 Baseline coverage, objectives, and adapters

Every acquisition family in Table 5 has a row in Tables 2, 14 and 26. The comparisons include BAX, PS-BAX, MF-MES, DAD, cost-aware BO, BAMCP, and fixed-policy controls. The common task exposes four supported intervention programs and four feasible measurement types per decision round; masks can remove candidates. All adapters return a typed query or stop, then use the same Bayesian update and terminal decision rule. Their names identify methodological ancestry; they are not claims of reproducing published performance on the original tasks.

The BAX target is a computable downstream property, whereas EIG targets the complete state (Neiswanger et al., 2021). PS-BAX uses a posterior-sampled target set (Cheng et al., 2024a); mapping that set to indirect biological assays is an additional adapter choice. MF-MES (Takeno et al., 2020) and cost-aware BO (Xie et al., 2024) need an explicit link from assay outcomes to intervention loss. The adapters retain the common assay menu; their original direct-function-observation guarantees do not automatically transfer. Also report native direct-evaluation subsets separately if those channels are available, without pooling their easier observation setting into the main comparison.

A.9.2 Resource matching and planner optimization

The online cap is $N = 4096$ conditional simulator calls per decision round, with lookahead depth two, 64 joint posterior particles, and a common bank of 64 predictive trajectories per intervention. A simulator call generates one conditional transition/retained-outcome sample; one batched call is charged by its sample count. Charge nested continuation draws and Step-DAD refinement to the same cap. Report likelihood evaluations, cached versus fresh trajectory draws, and realized calls separately: equal call caps do not imply equal wall time. Acquisition cost budget five measures scientific resources and is distinct from this computational cap. Fixed rules and amortized policies may use fewer online calls; unused capacity is reported, not padded. The same four-query/four-intervention workload and one 48-GiB GPU/eight host cores define the timing comparison.

For DAD, iDAD, Step-DAD, and amortized decision-aware BED, the additional policy-training cap is 2×10^6 simulator calls and 100,000 gradient updates per seed, with a cap of two million trainable policy parameters and seeds 40–44. Tune hyperparameters on disjoint validation groups within that allowance. Record training time and memory separately from the common posterior/WAM training; non-amortized methods receive no additional policy-training data. Existing resource profiles in Table 33 use a different serving workload and are not substitutes for these measurements.

Uniform versus adaptive allocation. Both numerical planners optimize the same depth-two decision-value target and include stop with value zero. Uniform BED cycles through every feasible root query in equal-size nested-MC batches. The NeuronSifter optimizer uses the same estimator but redistributes batches toward unresolved near-best queries. Allocate the first half of the cap uniformly across feasible roots: at the default cap and $K = 4$, each root receives a 512-call pilot allowance. Use 16 inner posterior particles per nested estimate, reweighted with the common retained-outcome likelihood. Complete an estimate only when all of its conditional draws and depth-two continuation

Table 35: Acquisition objectives and interface adapters. All use the same posterior, observation channel, action/query support, terminal loss, cost units, feasibility masks, and stopping options. The objective and its numerical optimizer are the controlled changes.

Method	Target and selection rule	Interface adaptation / limitation
EIG; BALD	Information about the joint latent state; information about the discrete mechanism regime.	Retained-outcome likelihood includes failure, censoring, and admission events.
BAX	Information about $A(f) = \arg\min_a f(a)$, with $f(a)$ the sampled intervention loss.	Integrate $I(A(f); \widetilde{o}_q \mid b)$ through the common assay channel; cost-normalize the score.
PS-BAX	Draw one posterior loss function and execute A to obtain a sampled minimizing action.	Choose the assay with greatest expected variance reduction for that action's loss per cost. This is an indirect-observation adapter.
MF-MES	Information about $f^* = \min_a f(a)$ per expected cost.	Register fidelity-specific channels and their correlations on training data; do not treat missing assays as noiseless low-fidelity evaluations.
DAD; iDAD; Step-DAD	Explicit-likelihood amortized information gain; implicit information gain; online-refined design.	Train finite-query logits with identical histories and masks; retain each method's objective and account for Step-DAD refinement calls.
Decision-aware BED	Expected terminal-loss reduction net of acquisition cost.	Separate the amortized head from the uniform nested-MC and adaptive-allocation optimizers.
Cost-aware BO	Pandora-style index for the assay-induced improvement of the incumbent intervention.	Form marginal improvement distributions through posterior predictive assay outcomes; the marginal index approximation ignores cross-query dependence.
BAMCP	Root-sampled tree search for negative terminal loss minus query costs.	Sample a latent model at each root rollout; preserve it down the path. Use common depth and predeclared observation bins.
Fixed controls	Random, cheapest, highest-fidelity, and fixed cascade.	Choose only feasible queries. Cascade order is exposure, engagement, clearance, function; skipped/unavailable calls are recorded.

evaluations fit the remaining call allowance; otherwise retain the unused remainder. The pilot allowance scales with the cap in the convergence sweep. At each allocation step let

$$U_q = \widehat{v}_q + 2s_q/\sqrt{n_q}, \qquad L_q = \widehat{v}_q - 2s_q/\sqrt{n_q}.$$

Maintain roots with $U_q \geq \max(0, \max_j L_j)$ and allocate the next affordable batch to the retained root with largest estimated standard error. Common random numbers pair candidate actions within an estimate, and pilot/completion rules and ties are fixed on validation data. These bands guide computation; they are not certified confidence bounds. At the cap, choose the largest estimated net value, or stop if it is nonpositive. Stop semantics, loss, posterior, and feasible support remain identical. The risk difference $0.166 - 0.160 = 0.006$ is an implementation-level contrast; causal attribution to the allocation rule requires paired runs that vary only that rule.

BAMCP (Guez et al., 2012) uses the same depth and simulator-call cap, UCT exploration coefficient $\sqrt{2}$, eight observation bins per continuous channel fixed on training data, and separate failure/status branches. Root sampling hides the drawn parameters from the policy; no privileged environment state enters a tree decision. Stop is available at every node, and terminal rollout value uses the same posterior loss. Log the number of completed simulations, node expansions, bin merges, and remaining unspent calls. The finite discretization and search cap are approximation choices that must be varied in sensitivity analysis.

At increasing budgets, compare terminal risk and root-query optimization gap against a separate 262144-call reference on the same fixed beliefs. This high-budget reference is an approximation, not an oracle; verify stability at twice its cap on a validation subset. Its costs are evaluation overhead, not free policy computation. The values in Table 37 describe budget sensitivity of the optimization gap,

Table 36: **Computational matching and resource use.** Caps are protocol settings. Online caps count conditional simulator calls per decision; offline calls are additional policy-training calls per seed. Offline training time and peak memory are distinct from online costs. Policies without online simulation still incur inference/likelihood costs.

Method	Online cap	Offline cap	Online (s)	GPU (GiB)
NeuronSifter	4096	0	10.80	28.8
Matched numerical BED	4096	0	9.60	28.4
BAMCP adapter (Guez et al., 2012)	4096	0	14.20	29.1
Decision-aware amortized BED	0	2×10^6	0.12	27.0
BAX (Neiswanger et al., 2021)	4096	0	12.90	28.9
PS-BAX adapter (Cheng et al., 2024a)	4096	0	3.40	27.4
MF-MES adapter (Takeno et al., 2020)	4096	0	5.10	27.8
Cost-aware BO adapter (Xie et al., 2024)	4096	0	4.80	27.7
EIG (nested MC)	4096	0	9.90	28.5
BALD (regime target)	4096	0	3.00	27.2
DAD (Foster et al., 2021)	0	2×10^6	0.10	27.0
iDAD adapter	0	2×10^6	0.12	27.0
Step-DAD adapter	4096	2×10^6	2.60	27.7

Table 37: **Computational budget sensitivity.** The first row defines a lower-cap sensitivity; the default cap is 4096. Root-query gap uses the separate high-budget reference on common beliefs. Timing excludes common WAM rollout generation and offline training.

Call cap	Ours risk	BED risk	BAMCP risk	Ours gap	BED gap	Ours (s)	BED (s)
1024	0.174	0.184	0.183	0.014	0.024	2.8	2.4
4096	0.160	0.166	0.169	0.007	0.013	10.8	9.6
16384	0.156	0.158	0.160	0.003	0.005	42.1	38.6
65536	0.154	0.155	0.156	0.001	0.002	168.4	154.2

including approximate parity at the largest budget. This finite-budget comparison does not establish asymptotic convergence. A wall-time-matched sensitivity run is also required because tree search, density evaluation, and cached rollout reuse have different costs.

A.9.3 Crossed independently generated environments

The oracle channel in Table 14 supplies privileged information inside the evaluation model; it tests information headroom, not whether that model is faithful. Holding out a registered teacher identity or changing its parameters still shares assumptions with training. The external test therefore separates the deployed twin from both the intervention generator and the observation generator, and freezes posterior weights, WAM, compiler, policy, and calibrator before accessing either generator.

The independent intervention-generator protocol specifies a separate codebase with a finite-volume transport solver, kinetic binding instead of the compiler's equilibrium shortcut, saturating clearance, and delayed inflammatory feedback. Fit or specify its parameter distributions independently of WAM training records; record their provenance and overlap rather than reusing teacher draws. It receives physical dose/route/schedule programs directly, never the twin's compiled occupancy paths.

The independent observation-generator protocol specifies its own mixing/saturation map, correlated measurement errors, and state-dependent censoring/admission. The policy sees only the resulting typed readouts and uses its frozen inference model, which may be misspecified.

Cross the registered versus independent transition with the registered versus independent observation generator. Use 64 independent contexts per cell and five model seeds; pair contexts and exogenous conditional outcome randomness across methods within each cell. Generate query outcomes on demand instead of replaying training outcomes. For path-dependent or invasive measurements, branch the environment from the current state rather than assuming a static outcome matrix. Use separate generator seeds and parameter draws for training, calibration, and evaluation. Before execution, retain hashes of the two generator implementations/configurations, a disjointness manifest, and the freeze time.

Terminal risk is scored using realized loss in the evaluation environment for the chosen intervention, including declared deferral/failure loss, never by asking the WAM to grade its own action. An environment-only evaluator computes the best feasible action after policy completion to measure regret; the policy cannot access that counterfactual panel. Report one-step and multistep predictive CRPS, intervention ordering, accepted fraction, false acceptance, and frozen-bound coverage along-

Table 38: **Independent-environment evaluation.** Risk is evaluated in the designated environment at scientific cost budget five. Coverage is for the frozen nominal-0.95 upper-loss bound. The oracle is specific to each environment and hidden from every policy.

Transition / observation	CRPS	Ours risk	BED risk	BAMCP risk	EIG risk	Oracle	Coverage
Registered / registered	0.110	0.160	0.166	0.169	0.193	0.110	0.953
Registered / independent	0.128	0.181	0.184	0.188	0.210	0.126	0.922
Independent / registered	0.157	0.197	0.201	0.198	0.226	0.142	0.875
Independent / independent	0.182	0.218	0.219	0.216	0.241	0.160	0.844

Table 39: **External diagnostics.** Posterior-state NRMSE uses latent truth available only to the evaluator. Ordering uses the external intervention panel. Regret is excess loss over that environment's oracle; false acceptance uses all contexts as denominator.

Transition / observation	Risk [95%]	State error	Ordering	Regret	Accepted	False accept
Registered / registered	0.160 [0.153, 0.167]	0.148	0.880	0.050	0.812	0.018
Registered / independent	0.181 [0.171, 0.191]	0.171	0.844	0.055	0.781	0.031
Independent / registered	0.197 [0.184, 0.210]	0.196	0.812	0.055	0.719	0.047
Independent / independent	0.218 [0.204, 0.232]	0.227	0.766	0.058	0.625	0.078

side risk. Independent transition and observation changes degrade both prediction and decisions, while BAMCP has slightly lower risk in the joint-shift cell. Independent simulator tests assess robustness to specified model discrepancies, not biological validity.

A.9.4 Four separate sources of decision value

Table 40: **Four controlled contrasts.** A/B give reference and changed condition; external means independently generated transition and observation. $\Delta = R_B - R_A$. These conditional effects are not additive.

Axis: A versus B	Replay risk A/B	External risk A/B	External Δ [95%]
Posterior: joint / factorized	0.160/0.199	0.218/0.252	0.034 [0.030, 0.038]
Compiler: occupancy / action ID	0.160/0.218	0.218/0.267	0.049 [0.044, 0.054]
Objective: decision value / EIG (uniform MC)	0.166/0.193	0.219/0.241	0.022 [0.016, 0.028]
Planner: adaptive / uniform allocation	0.160/0.166	0.218/0.219	0.001 [-0.004, 0.006]

The posterior comparison replaces joint particles with a factorized approximation whose univariate marginals are retained, while keeping occupancy compilation and the adaptive planner. The compiler comparison replaces occupancy with action identity in a parameter-matched retrained WAM, keeping posterior and planner. This estimates the value of the complete conditioning interface, including retraining, rather than a runtime-only software change. For the acquisition-objective comparison, both EIG and decision-aware BED use uniform nested MC and identical stopping evaluation; only their query ranking objective differs. Finally, the planner comparison fixes posterior, compiler, WAM, loss, and decision-value objective and varies allocation alone. Calibrate each changed policy on the same disjoint groups before freezing; use identical external test contexts thereafter. The contrasts are conditional effects and must not be added as an additive decomposition; a factorial study is needed to estimate interactions.

The external-environment contrast for allocation is 0.001 with an interval spanning zero, whereas the other three contrasts are larger. An interval spanning zero does not support a resolved allocation effect or an architecture advantage. Paired episode-level estimation and interval reporting should cluster by independent context/source and keep training-seed variation separate. Existing point-state, action-identity, and information-objective ablations remain in Table 19; they do not replace this controlled external evaluation.

B NOTATION AND ASSUMPTIONS

Table 41: Core objects in the Agent-in-Twin formulation.

Symbol	Meaning
$\boldsymbol{\xi}_t = (\mathbf{u}_t, m, \boldsymbol{\theta}, \boldsymbol{\rho}_t)$	latent neuronal microenvironment, mechanism, parameters, and tool reliability
$b_t \in \mathcal{B}$	updateable digital-twin posterior
$\Phi : \mathcal{A}^{\mathrm{w}} \to \Omega$	structured world-action compiler
$\mathsf{P}_{\Phi(a)}$	WAM transition kernel under compiled action a
O_q	typed value/status/cost/provenance channel of query q
π^{AIT}	internal policy over world actions, information actions, and terminal decisions
$\Gamma(b_t)$	external or compressed posterior interface used by a comparison policy
$\mathcal{R}(b)$, $\mathrm{VOI}(q; b)$	terminal Bayes risk and decision-focused information value
$\mathcal{G}_{0:H}$	bounded cumulative decision loss over a closed-loop horizon

We use normalized total variation $\mathrm{TV}(P, Q) = \sup_A |P(A) - Q(A)|$. Assumptions are grouped by the conclusions they support.

Assumption 1 (Bounded losses)**.** Terminal loss satisfies $0 \le \ell(d, \boldsymbol{\xi}) \le L$. Closed-loop path loss satisfies $0 \le \mathcal{G}_{0:H} \le C_H$.

Assumption 2 (Typed action semantics)**.** Pure information actions leave biological state unchanged; invasive measurements are hybrid actions with explicit WAM components. World transition and query-outcome channels are separately registered.

Assumption 3 (Compiler fidelity)**.** For every raw program a and state $\boldsymbol{\xi}$, the physical transition $\mathsf{P}^{\mathrm{raw}}_a$ and compiled WAM transition satisfy $\mathrm{TV}(\mathsf{P}^{\mathrm{raw}}_a(\cdot \mid \boldsymbol{\xi}), \mathsf{P}_{\Phi(a)}(\cdot \mid \boldsymbol{\xi})) \le \epsilon_\Phi$. Exact sufficiency is the case $\epsilon_\Phi = 0$.

Assumption 4 (Closed-loop model approximation)**.** The true and learned initial biological-state laws differ by at most ϵ_b in total variation. At matched physical histories, WAM transitions and typed outcome channels differ by at most ϵ^{W}_t and ϵ^{O}_t, uniformly over the common feasible actions. A fixed deployed controller, including its posterior-update computation, is executed in both environments. Its action randomization can therefore be shared while observed histories agree. The path comprises biological states, executed actions, observations, statuses, costs, and terminal decisions; belief arrays are deterministic functions of the observed history.

Assumption 5 (Planning approximation)**.** The learned policy is ϵ_{plan}-optimal in the learned environment within Π_Γ, and this same policy class is available in both environments. The class includes the retained posterior representation and deployed update rule, as well as time, budget, and feasibility context.

Assumption 6 (Common query channel for coherence analysis)**.** When beliefs b and b' are compared for a fixed query q, they use the same registered channel $\mathsf{O}_q(d\mathbf{o} \mid \boldsymbol{\xi})$. Belief-dependent execution penalties are treated separately. For the nonnegative-VOI and local coherence results, T_q is Bayesian conditioning under this channel, including its retained-outcome likelihood, admission events, and modeled missingness. The learned amortized update is an approximation whose consistency is evaluated separately.

Assumption 7 (Utility approximation)**.** $|\widehat{\mathcal{U}}_\omega(b, q) - \mathcal{U}(b, q)| \le \epsilon$ uniformly over the feasible query set at the evaluated belief.

Assumption 8 (Selective calibration)**.** For the selected decision d_b, the calibration procedure supplies a bound satisfying $\mathbb{P}\{\ell(d_b, \boldsymbol{\xi}) > \widehat{U}_\delta(d_b, b)\} \le \delta$ under its stated assumptions. A uniform bound or direct calibration of the decision rule covers this data-dependent selection. Support and shift gates are frozen before test evaluation.

Algorithm 1 A NeuronSifter intervention-planning round

Require: Evidence $\mathcal{E}_t$, microenvironment graph $\mathbf{G}$, interventions $\mathcal{W}$, queries $\mathcal{Q}$, budget B_t, evaluation mode

Infer the joint biological posterior b_t ▷ Retain exposure–response uncertainty
for $a \in \mathcal{W}$ **do**
    Compile supported exposure and occupancy programs for posterior states
    Sample CNS trajectories and estimate loss in Equation (7)
Rank interventions and compute TRI, support, and ranking stability
if fixed-risk mode and the frozen stopping rule returns a candidate **then**
    return candidate if the terminal calibrated gate accepts; otherwise defer
for $q \in \mathcal{Q}(b_t, B_t)$ **do**
    Predict typed outcomes and their posterior intervention-risk reduction
    Subtract expected measurement cost and failure penalty; enforce the latency limit
Compare feasible continuations with stopping under the decision objective
if stopping is preferred or no feasible budget remains **then**
    return the best admissible decision or an evidence-based deferral
Select the preferred feasible information/world action
if an action is executed **then**
    Propagate any world-action component and obtain its typed observation
    Assimilate retained fields and typed events into b_{t+1} using Equation (10)
    Update remaining budget using realized cost ▷ Count partial and failed calls
Repeat until a decision is resolved or the budget is exhausted

C Extended Agent-in-Twin Method

C.1 Closed generative–control factorization

Condition on the initial evidence $\mathcal{E}_0$. A physical trajectory and its acquired observations factorize as

$$\begin{aligned} &p(\boldsymbol{\xi}_{0:T}, a^{\mathrm{w}}_{0:T-1}, a^{\mathrm{i}}_{0:T-1}, \mathbf{o}_{1:T}, d_T \mid \mathcal{E}_0, \mathbf{G}) \\ &\quad = p(\boldsymbol{\xi}_0 \mid \mathcal{E}_0, \mathbf{G}) \prod_{t=0}^{T-1} \Big[\pi_t^{\mathrm{AIT}}(a^{\mathrm{w}}_t, a^{\mathrm{i}}_t \mid b_t, B_t) \mathsf{P}_{\Phi(a^{\mathrm{w}}_t)}(\boldsymbol{\xi}_{t+1} \mid \boldsymbol{\xi}_t, \mathbf{G}) \\ &\qquad\qquad \times \mathsf{O}_{a^{\mathrm{i}}_t}(\mathbf{o}_{t+1} \mid \boldsymbol{\xi}_{t+1}) \Big] \pi_T^{\mathrm{AIT}}(d_T \mid b_T, B_T). \end{aligned} \tag{15}$$

b_t is computed from observed history. The observation channel includes passive measurements; early stopping is absorbing. World and information actions respectively control state transitions and observation.

C.2 Belief representation and uncertainty decomposition

The distributional reference uses a finite-support categorical variational family

$$q_\phi(d\xi \mid \mathcal{E}) = \sum_{j=1}^{J} w_{\phi j}(\mathcal{E})\, \delta_{\xi^{(j)}}(d\xi), \quad w_{\phi j} = \mathrm{softmax}_j\, s_\phi(\mathcal{E}, \xi^{(j)}), \tag{16}$$

Each tuple jointly contains treatment history, PK/exposure, environment, electrophysiology, morphology, and regime, preserving constraints and dependence. Shared finite prior/variational support gives categorical KL in Equation (26). Amortized weights update as $w_j^+ \propto w_j \overline{p}_q(\widetilde{o} \mid \xi^{(j)})$. Normalize/resample whole tuples; report particle count, effective sample size, and support failures.

The joint particles in Equation (16) support observations, intervention comparison, and acquisition (Figure 2). Treatment history is observed context; proposed interventions enter the compiler. Heads are conditionally independent given joint tuples and source effects. Linked assays retain sample identity; held-source residuals test remaining dependence. Calibrated joint sampling remains an evaluation target beyond deterministic feature heads.

RNA counts use negative-binomial likelihoods with library-size offsets; positive concentrations use log-normal likelihoods with detection-limit censoring; morphology/electrophysiology use heteroscedastic Gaussians after training-only scaling; subtype/state evidence uses categorical likelihoods. Direct occupancy requires the registered assay law and its endpoint/censoring rules.

Report modality/source-wise NLL, continuous CRPS or categorical Brier score, 50/90/95% coverage, PIT (randomized for counts), and classwise reliability with denominators. Retain missingness/support and cross-modal residuals before aggregation. Calibration uses disjoint groups; observation-head and terminal-decision calibration are separate.

For a scalar trajectory functional $h(\boldsymbol{\tau})$,

$$\begin{aligned}\mathrm{Var}[h(\boldsymbol{\tau}) \mid b, a] = \mathbb{E}_{\boldsymbol{\xi}\sim b}[\mathrm{Var}(h(\boldsymbol{\tau}) \mid \boldsymbol{\xi}, a)] \\ + \mathrm{Var}_{\boldsymbol{\xi}\sim b}(\mathbb{E}[h(\boldsymbol{\tau}) \mid \boldsymbol{\xi}, a]),\end{aligned} \tag{17}$$

separating conditional rollout variation from posterior uncertainty. The acquisition target measures how their joint uncertainty affects intervention loss.

C.3 Structured action compiler

For pharmacological actions, the compiler uses

$$\omega_{g,h}^{\mathrm{target}} = \frac{C_{g,h}^{\mathrm{free}}}{C_{g,h}^{\mathrm{free}} + K_{d,g}}, \qquad C_{g,0:H}^{\mathrm{free}} = \mathrm{PK}_{\nu}(d, \varrho, s, \Delta, \chi), \tag{18}$$

with posterior uncertainty over PK, free concentration, and affinity. The compiler schema is

$$\Phi(a, \boldsymbol{\xi}, \mathbf{G}) = (\omega^{\mathrm{target}}, \omega^{\mathrm{exposure}}, \omega^{\mathrm{stim}}, \omega^{\mathrm{environment}}, \omega^{\mathrm{residual}}, M^{\mathrm{support}})_{0:H}. \tag{19}$$

The compiled object is the state-conditional map with cost and support; equivalence holds over reachable states. The ratio assumes equilibrium binding. Registered kinetic models cover nonequilibrium mechanisms while retaining dose, route, schedule, and CNS context.

C.4 Occupancy-conditioned diffusion operator

Trajectories have sample, intervention, time, region, compartment, and species axes. With $N = RC$, a three-level temporal transform packs $Y = X - X_0$ into coarse/detail coefficients; odd lengths use replicated boundaries and are cropped on inversion. The tokenizer uses the symmetric graph Laplacian $L = D - (A + A^{\top})/2 = U\Lambda U^{\top}$ to supply spectral context. This feature graph is distinct from the directed transport model used by the action compiler.

Let $c = \mathcal{H}(Y)$ and v_k denote codebook values. Coefficients are clipped to the declared range and quantized to tokens k_j. Following absorbing masked diffusion (Sahoo et al., 2024), each token is replaced by a mask independently with probability $u \in (0, 1)$. The denoiser predicts its clean categorical distribution conditional on the corrupted field and context C. The masked training objective is

$$\mathcal{L}_{\mathrm{mask}} = \mathbb{E}_{u,M}\left[-\frac{1}{Du}\sum_{j=1}^{D} M_j \log p_{\theta}(k_j \mid c_u, C, u)\right], \tag{20}$$

where D is the number of coefficients. Nonempty-mask sampling uses its probability mass as an importance correction. Reconstruction compares the inverse transform with an independent teacher, retaining unmasked coefficients and using predicted codebook expectations at masked positions. Teacher confidence weights both losses.

Conditioning supplies aligned occupancy/support, initial state, posterior context, and physical/diffusion time. Blocks combine self-attention, context cross-attention, semantic mixing, ordered experts (Equation (6)), and feed-forward updates. AdaLN-Zero modulates tokens; expert-scope annotations guide routing.

The expert composition in Equation (6) uses $h^{(0)}$ as the incoming representation and passes each expert's updated features to the next. The parallel control evaluates all experts from the same incoming representation.

Reverse sampling starts masked. Each step $u \to v < u$ reveals masked tokens with probability $(u - v)/u$ and samples categorical predictions; revealed tokens stay fixed and the last step resolves all masks. Inversion adds X_0, preserving the initial condition. Repeated draws give predictive summaries; continuous residual refinement is a separate ablation.

Action contrasts. The paired-difference objective is defined on source-supported programs a, a' sharing one context: $\mathcal{L}_{\mathrm{contrast}} = \mathbb{E}\|M \odot [(\widehat{X}^{a} - \widehat{X}^{a'}) - (X^{a} - X^{a'})]\|_2^2/\|M\|_1$, where M retains jointly observed channels and time points. Pairs with empty M are excluded. No-action, exposure-matched, mechanism-matched, and sentinel programs isolate different intervention components.

C.5 Decision-focused acquisition

The exact post-query risk is

$$\mathcal{R}^q(b,q) = \mathbb{E}_{\mathbf{o}\sim p(\cdot|b,q)}\left[\min_{d\in\mathcal{D}_{\mathrm{dec}}} \mathbb{E}_{\boldsymbol{\xi}\sim \mathsf{T}_q(b,\mathbf{o})}\ell(d,\boldsymbol{\xi})\right]. \tag{21}$$

Nested Monte Carlo uses the execution-time admission and update code. Logged-policy learning and comparisons require positive historical action support; learning uses clipped inverse propensity weights and reports effective sample size. Full-outcome matrices supply offline comparisons across the feasible query library.

C.6 Multi-step internal planning

With k remaining queries and budget B,

$$\begin{aligned} V_0(b,B) &= \mathcal{R}(b), \\ V_k(b,B) &= \min\left\{\mathcal{R}(b), \min_{q\in\mathcal{Q}(b,B)} \lambda\bar{\mathsf{c}}(b,q) + \mathbb{E}_{\mathbf{o}}[V_{k-1}(\mathsf{T}_q(b,\mathbf{o}), B-\mathsf{c}(\mathbf{o}))]\right\}. \end{aligned} \tag{22}$$

The lookahead target is attainable continuation loss reduction, with $W_0(b,B)=0$ and recursion

$$W_k(b,B) = \max\left\{0, \max_{q\in\mathcal{Q}(b,B)} [\mathcal{U}(b,q) + \mathbb{E}_{\mathbf{o}} W_{k-1}(\mathsf{T}_q(b,\mathbf{o}), B-\mathsf{c}(\mathbf{o}))]\right\}. \tag{23}$$

Stopping/empty feasible sets have value zero. Nested Monte Carlo on out-of-fold posteriors uses the execution channel to form $\widetilde{W}_k$; fit $\mathcal{L}_{\mathrm{look}} = \mathbb{E}[(\widehat{W}_\zeta - \widetilde{W}_k)^2]$ conditional on rounds/budget. Targets use a frozen previous head or complete bounded lookahead. Held-out episodes test utility error. Deploy one-step net utility plus expected continuation value with unit discount; world actions follow Equation (28).

C.7 Reliability-aware assimilation

The coarsened outcome (Equation (10)) retains admitted fields and completion/admission events. State-dependent rejection contributes its event likelihood, including detection-limit tails. If execution status depends only on tool reliability and rejected-field admission is ignorable given retained metadata, the likelihood is $p(\zeta, \mathsf{c}, \mathrm{latency} \mid \boldsymbol{\rho}, q)\, p(y_{\mathrm{adm}}, \sigma_{\mathrm{adm}} \mid \boldsymbol{\xi}^{\mathrm{bio}}, q, \zeta)$. Fixed admission rules check identity, source role, units, context, uncertainty, and query-specific QC. Partial records integrate missing fields; below-detection records retain censoring. Timeouts retain costs and update tool reliability. Invalid schema/units, unsupported contexts, and protected roles quarantine values without converting them to zero or therapeutic actions. Rejections update biology only through validated state-dependent event likelihoods; otherwise biological assimilation is withheld with insufficient-evidence status. Valid contradictory records retain their source likelihoods. Held-source predictive and decision calibration evaluate learned updates to the retained channel.

Selective intervention states. Accept/reject require calibrated risk and support; active learning requires positive net value within budget. External validation handles decisions sensitive to evidence beyond registered tools. Unsupported decisions return insufficient evidence. Report coverage and cost for each outcome.

C.8 TRI and selected-decision calibration

TRI combines normalized calibration, evidence coverage, PK/PD consistency, ranking stability, and in-distribution support with nonnegative weights:

$$\mathrm{TRI}(b,a) = \frac{\sum_{j=1}^5 w_j r_j(b,a)}{\sum_{j=1}^5 w_j}. \tag{24}$$

The configured weights are $(0.25, 0.20, 0.20, 0.20, 0.15)$; all components increase with reliability. Hard feasibility and support predicates remain separate.

Freeze policy, acquisition, risk prediction, and stopping before calibration. Execute the frozen rule on n exchangeable episodes to obtain $e_i = \ell(d_i,\boldsymbol{\xi}_i) - \widehat{r}(d_i,b_i)$. Let q_δ be the $\lceil (n+1)(1-\delta)\rceil$ order statistic, using $+\infty$ if this index exceeds n. A split-calibrated upper bound is

$$\widehat{U}_\delta(d,b) = \min\{L, \max\{0, \widehat{r}(d,b) + q_\delta\}\}. \tag{25}$$

Exchangeability gives marginal coverage including adaptive query histories. Shared sources require independent group calibration; new policies or threshold families require recalibration.
After stopping, accept/defer uses the bound with $\mathrm{TRI} \geq \tau_r$, feasibility, and support; acquisition remains closed. Unconditional false-acceptance control is preserved, while conditional risk retains the coverage factor (Theorem 6). Evaluate shifts separately.

C.9 Training objectives and cross-fitting

With $\boldsymbol{\xi} = (\mathbf{u}, m, \boldsymbol{\theta}, \boldsymbol{\rho})$, the twin is trained by

$$\begin{aligned}\mathcal{L}_{\text{belief}} = & -\mathbb{E}_{q_\phi(\boldsymbol{\xi}|\mathcal{E})}[\log p_\psi(\mathbf{x} \mid \boldsymbol{\xi}, \mathbf{G})] + \beta \mathrm{KL}(q_\phi(\boldsymbol{\xi} \mid \mathcal{E}) \| p(\boldsymbol{\xi} \mid \mathbf{G})) \\ & + \lambda_{\text{hold}} \mathcal{L}_{\text{held-modality}} + \lambda_{\text{ont}} \mathcal{L}_{\text{ontology}} + \lambda_{\text{src}} \mathcal{L}_{\text{source-invariance}}. \end{aligned} \tag{26}$$

The source-nuisance branch and held-source calibration are shared with the main-text posterior interface.
The WAM objective couples trajectory likelihood to decision-relevant action structure,

$$\mathcal{L}_{\text{WAM}} = \mathcal{L}_{\text{traj}} + \lambda_{\text{score}} \mathcal{L}_{\text{CRPS}} + \lambda_{\text{act}} \mathcal{L}_{\text{contrast}} + \lambda_{\text{phys}} \mathcal{L}_{\text{constraint}} + \lambda_{\text{repr}} \mathcal{L}_{\text{distill}}. \tag{27}$$

Action contrasts use the source-supported pairs defined in Appendix C.4.
The amortized heads approximate retained-channel Monte Carlo with score

$$\widehat{Q}(b_t, B_t, a_t) = \widehat{r}_{\text{dec}}(b_t, a_t) + \gamma \mathbb{E}_{\widetilde{\mathbf{o}}_t}[\widehat{V}_\zeta(\mathsf{T}(b_t, a_t, \widetilde{\mathbf{o}}_t), B_t - \widetilde{\mathsf{c}}_t)], \tag{28}$$

Here $\widehat{r}_{\text{dec}}$ is immediate net loss reduction, $\widehat{V}_\zeta$ estimates attainable continuation value, and $\gamma = 1$ for the undiscounted objective.
The update first coarsens the outcome. The learning objectives match this retained evidence and its decision value:

$$\mathcal{L}_{\text{acq}} = \mathbb{E}[(\widehat{\mathcal{U}}_\omega(b, q) - \widetilde{\mathcal{U}}_{\text{MC}}(b, q))^2] + \lambda_r \mathcal{L}_{\text{rank}} + \lambda_v \mathcal{L}_{\text{look}}, \tag{29}$$

$$\mathcal{L}_{\text{update}} = \mathbb{E}[\mathrm{KL}(\mathsf{T}(b, q, \mathbf{o}) \,\|\, q_\phi(\boldsymbol{\xi} \mid \mathcal{E} \cup \{\widetilde{\mathbf{o}}\}, \mathbf{G}))]\,. \tag{30}$$

Combining these objectives gives

$$\mathcal{L}_{\text{AIT}} = \mathcal{L}_{\text{belief}} + \alpha_w \mathcal{L}_{\text{WAM}} + \alpha_q \mathcal{L}_{\text{acq}} + \alpha_u \mathcal{L}_{\text{update}} + \alpha_s \mathcal{L}_{\text{selective}}. \tag{31}$$

Fit belief/WAM models on admitted training groups; use out-of-fold posteriors/outcomes for acquisition and update consistency, disjoint calibration groups for thresholds, and untouched source/action/simulator/tool holdouts.

C.10 Stability and identifiability under partial observation

The temporal transform preserves the initial state on inversion, while codebook coefficients constrain decoded magnitudes. Expert residual scales and AdaLN-Zero initialization control update size. If routing coefficients are externally fixed and expert e is L_e-Lipschitz, the ordered residual bank has sensitivity at most $\prod_e (1 + |\alpha_e \pi_e| L_e)$. End-to-end sensitivity additionally includes derivatives of π_e; contraction requires further bounds on the complete update. Nonfinite outputs, physical-constraint violations, and finite-difference gain at one, two, and four times the reporting horizon provide rollout diagnostics.
Decision identifiability compares observationally compatible states within a validation-fixed held-observation log-score tolerance. Under common noise, structurally clamp exposure, inflammation, or transport and recompute posterior loss. Report signed contrasts, compatible-scenario ranges, and rank-change frequency; sign-changing ranges mark ambiguity. Masks, tolerances, and clamps match across operators. Biological causal identification additionally requires independent intervention evidence.

D Theoretical Results

We analyze intervention compilation, retained biological information, and closed-loop error using action abstraction and belief-state control (Ravindran & Barto, 2004; Kaelbling et al., 1998).

D.1 Posterior interface and decision sufficiency

Proposition 1 (Posterior-preserving interface principle)**.** *Let an external policy observe $r_t = \Gamma(b_t)$ under the same actions, budgets, kernels, and loss. The optimal Agent-in-Twin value is no larger than the optimal value restricted to r_t. Predictive sufficiency for the relevant transition, observation, loss, and update kernels is sufficient for equality. If an interface merges beliefs requiring different optimal continuations, a decision problem separates the two policy classes.*

Controller placement does not enter this statement. With the full posterior, update rule, action support, time/budget context, costs, loss, and policy class matched, either controller can emulate the other. The full-posterior external control tests this equality; compressed interfaces test which information determines the CNS decision. Finite-compute optimization error is a separate comparison.

D.2 Structured world-action sufficiency

Theorem 1 (Exact WAM action sufficiency)**.** *For a finite intervention library, suppose $\epsilon_\Phi = 0$ in assumption 3, compiled actions retain feasibility and declared cost, and path loss depends on a program through its induced trajectory and this cost. The feasible compiled set is the image of the state-conditional compiler. Then optimizing over raw intervention programs and optimizing over their compiled WAM fields have the same finite-horizon optimal value. Any two WAM-equivalent programs are decision-equivalent.*

State-conditional compilation is a sufficient action abstraction when it preserves transitions, feasibility, and cost (Ravindran & Barto, 2004). The approximate result quantifies departures from this condition.

Theorem 2 (Approximate WAM value preservation)**.** *If path loss is bounded by C_H and the compiled transition differs from the raw-program transition by at most $\epsilon_{\Phi,t}$ in total variation at round t, then*

$$|J_{\mathrm{raw}}(\pi) - J_{\mathrm{WAM}}(\pi)| \le C_H \sum_{t=0}^{H-1} \epsilon_{\Phi,t}, \tag{32}$$

for corresponding policies with common initial law, observation channels, feasibility, and cost. The same bound holds for optimal values in the common policy class.

D.3 Finite-horizon intervention regret

For a common deployed controller, let ϵ_b bound initial biological-state error and $\epsilon_t^{\mathrm{W}}, \epsilon_t^{\mathrm{O}}$ bound transition and typed-observation errors at matched physical histories. Set

$$\Delta_H = \min\left\{1, \epsilon_b + \sum_{t=0}^{H-1} (\epsilon_t^{\mathrm{W}} + \epsilon_t^{\mathrm{O}})\right\}. \tag{33}$$

Let Π_Γ be the common policy class implementable through the retained posterior interface and update rule. Its approximation gap is $g_\Gamma = \inf_{\pi\in\Pi_\Gamma} J_{\mathcal{M}}(\pi) - \inf_\pi J_{\mathcal{M}}(\pi)$.

Theorem 3 (Closed-loop intervention regret)**.** *Under the common-policy and bounded-path assumptions in Appendix B, an ϵ_{plan}-optimal learned policy in Π_Γ satisfies*

$$J_{\mathcal{M}}(\widehat{\pi}) - \inf_\pi J_{\mathcal{M}}(\pi) \le 2C_H\Delta_H + \epsilon_{\mathrm{plan}} + g_\Gamma. \tag{34}$$

The error budget in Equation (33) compares physical path laws under the same deployed controller. The controller's update computation is held fixed in this comparison; approximation of the information it retains is represented by g_Γ.

Lemma 1 (Physical-path comparison)**.** *Under assumption 4, every fixed policy in the common class satisfies*

$$\mathrm{TV}(\mathbb{P}^\pi_{\mathcal{M}}, \mathbb{P}^\pi_{\widehat{\mathcal{M}}}) \le \Delta_H. \tag{35}$$

Couple the initial physical states and each transition/outcome at matched histories. Because the same controller sees the same history until a mismatch, its next action can be coupled identically. Summing first-mismatch probabilities proves the bound. Bounded path loss and learned-policy optimality then give Theorem 3.

Corollary 1 (Equivalent posterior interfaces)**.** *An external controller with the same posterior, update computation, time/budget context, feasible actions, and policy class has the same attainable decision values as Agent-in-Twin. An information-losing interface can increase g_Γ.*

The matched comparison isolates the decision value of retained biological information.

D.4 Decision-theoretic foundations and local coherence

Proposition 2 (Belief-state sufficiency)**.** *Under Equation* (15)*, the exact posterior β_t is sufficient for future typed outcomes and optimal policies with latent-state-dependent loss. Exact policies may be written as functions of (β_t, B_t) Kaelbling et al. (1998).*

Proposition 3 (Nonnegative value of information)**.** *Under assumption 6,* $\mathrm{VOI}(q; b) \ge 0$ *when the post-query rule may ignore the observation, the decision set is unchanged, and query cost is excluded.*

Theorem 4 (Belief-coherence stability)**.** *Under assumptions 1 and 6, for fixed q,*

$$|\mathcal{R}(b) - \mathcal{R}(b')| \le L\mathrm{TV}(b, b'), \tag{36}$$

$$|\mathcal{R}^q(b, q) - \mathcal{R}^q(b', q)| \le L\mathrm{TV}(b, b'), \tag{37}$$

$$|\,\mathrm{VOI}(q; b) - \mathrm{VOI}(q; b')| \le 2L\mathrm{TV}(b, b'). \tag{38}$$

Corollary 2 (Query-order stability)**.** *If Δ_{pen} bounds each query's penalty change, $q_1 \succ q_2$ under b remains ordered under b' whenever*

$$\mathcal{U}(q_1; b) - \mathcal{U}(q_2; b) > 4L\mathrm{TV}(b, b') + 2\Delta_{\mathrm{pen}}. \tag{39}$$

Theorem 5 (One-step acquisition regret)**.** *Under assumption 7, if $q^\star$ maximizes true utility and $\widehat{q}$ maximizes estimated utility, then $\mathcal{U}(b, q^\star) - \mathcal{U}(b, \widehat{q}) \le 2\epsilon$.*

D.5 Selective intervention control

Theorem 6 (False-acceptance control)**.** *Under assumption 8, define $A_\tau = \mathbf{1}\{\widehat{U}_\delta(d_b, b) \le \tau\}\mathsf{G}_{\mathrm{support}}(b)$. Then*

$$\mathbb{P}(\ell(d_b, \boldsymbol{\xi}) > \tau, A_\tau = 1) \le \delta. \tag{40}$$

Conformal risk control calibrates expected risks of specified decision-rule families under its own assumptions (Angelopoulos et al., 2024). The split-calibrated construction in Appendix C.8 targets the selected-decision guarantee in assumption 8 under exchangeable episodes.

Corollary 3 (Conditional accepted risk)**.** *If $\mathbb{P}(A_\tau = 1) \ge \kappa_{\mathrm{cov}} > 0$, then $\mathbb{P}(\ell > \tau \mid A_\tau = 1) \le \min\{1, \delta/\kappa_{\mathrm{cov}}\}$.*

D.6 Policy-specific path divergence

Policy-specific divergence weights model discrepancy by the histories visited during deployment. Let H_t contain the complete state and observation history needed by the kernels, and let h_t be the observable history available to the controller. Both models execute the same mapping $\pi(a_t \mid h_t)$. Stopping uses absorbing states up to a fixed horizon. Write

$$K_\pi = \mathrm{KL}(p_0 \| \widehat{p}_0) + \sum_{t=0}^{H-1} \mathbb{E}_{\mathbb{P}^\pi_{\mathcal{M}}} \Big[\mathrm{KL}(P_t(\cdot \mid \mathsf{H}_t, a_t) \| \widehat{P}_t(\cdot \mid \mathsf{H}_t, a_t)) + \mathbb{E}_{x' \sim P_t} \mathrm{KL}(\overline{O}_t(\cdot \mid x', \mathsf{H}_t, a_t) \| \widehat{\overline{O}}_t(\cdot \mid x', \mathsf{H}_t, a_t))\Big]. \tag{41}$$

The retained observation includes admission, failure, latency, and realized cost.

Proposition 4 (Policy-specific transfer)**.** *Suppose the augmented reference path law is absolutely continuous with respect to the learned path law. For common feasible actions and a common path loss $G \in [0, C_H]$, set $\delta_\pi = \min\{1, \sqrt{K_\pi/2}\}$. Then*

$$|J_{\mathcal{M}}(\pi) - J_{\widehat{\mathcal{M}}}(\pi)| \le C_H \delta_\pi. \tag{42}$$

If $\widehat{\pi}$ is ϵ_{plan}-optimal in a common class under the learned model, every comparator $\pi^\star$ in that class satisfies

$$J_{\mathcal{M}}(\widehat{\pi}) - J_{\mathcal{M}}(\pi^\star) \le C_H(\delta_{\widehat{\pi}} + \delta_{\pi^\star}) + \epsilon_{\mathrm{plan}}. \tag{43}$$

The chain rule gives augmented-path KL equal to K_π: shared policy factors cancel. Pinsker's inequality and bounded loss prove the first statement. Adding and subtracting the two learned-model values proves the second. Comparison with a larger policy class additionally incurs its expressivity gap. The two divergence terms concern the learned and comparator policies separately, each under its reference-model occupancy distribution.

What the diagnostics measure. Evaluable reference and learned conditional densities permit log-ratio estimates along a fixed policy. The KL estimate uses the ratio of both conditional densities. A single-model log score estimates cross-entropy. For the five-outcome reference model in Appendix A.1.8, categorical probabilities give exact TV and KL. Projection contracts TV, so categorical distance characterizes the declared outcome model. Discriminator and MMD scores provide complementary observation-level diagnostics; intervention regret tracks the corresponding decision loss.

E PROOFS

E.1 POSTERIOR-INTERFACE DOMINANCE

Let Π_b be the admissible policies measurable with respect to (b_t, B_t) and Π_Γ those measurable with respect to $(\Gamma(b_t), B_t)$. Composition with Γ gives $\Pi_\Gamma \subseteq \Pi_b$, hence $\inf_{\pi\in\Pi_b} J(\pi) \le \inf_{\pi\in\Pi_\Gamma} J(\pi)$. If Γ is predictively sufficient for every feasible typed kernel and loss, conditional future path laws depend on b_t only through $\Gamma(b_t)$; the standard state-abstraction construction maps an optimal posterior policy to an equally valued Γ-policy. If sufficiency fails, two posterior states share one interface value but induce different future outcome or loss laws. A bounded binary decision that chooses the preferred continuation separates them, proving strict inequality for some decision problem.

E.2 EXACT WAM ACTION SUFFICIENCY

The compiler partitions raw programs into equivalence classes. Under exact fidelity, every program in one class induces the same controlled kernel at every state. Since loss and cost depend on the program only through the induced trajectory and declared cost, replacing a raw program by its compiled representative leaves the law of every path and its loss unchanged. Mapping a raw policy through Φ and choosing any raw representative for a compiled policy therefore preserves value in both directions; taking infima proves equality of optimal values.

E.3 APPROXIMATE WAM VALUE PRESERVATION

Couple raw and compiled transitions at each round by a maximal coupling. Conditional on matched histories, the probability that the next states disagree is at most $\epsilon_{\Phi,t}$. A union bound over the horizon gives total path disagreement probability at most $\sum_t \epsilon_{\Phi,t}$. Because $0 \le \mathcal{G}_{0:H} \le C_H$, the difference in expected path loss is at most $C_H \sum_t \epsilon_{\Phi,t}$. Applying this result to optimal policies in each model and using the standard add-and-subtract argument yields the optimal-value statement.

E.4 PHYSICAL-PATH COMPARISON AND INTERVENTION REGRET

For Lemma 1, maximally couple the initial physical-state laws. Conditional on matched observed histories, execute the same deployed update and policy with shared randomization. Couple the biological transition and typed outcome channel with mismatch probabilities bounded by ϵ_t^{W} and ϵ_t^{O}. A union bound, capped at one, yields Equation (35). The coupling operates on physical histories while holding the controller computation fixed.

For every $\pi \in \Pi_\Gamma$, bounded loss implies

$$|J_{\mathcal{M}}(\pi) - J_{\widehat{\mathcal{M}}}(\pi)| \le C_H \Delta_H. \tag{44}$$

Let $\pi^\star_\Gamma$ minimize the true value within Π_Γ. Add and subtract learned values for $\widehat{\pi}$ and $\pi^\star_\Gamma$. Two applications of Equation (44) and ϵ_{plan}-optimality give

$$J_{\mathcal{M}}(\widehat{\pi}) - J_{\mathcal{M}}(\pi^\star_\Gamma) \le 2C_H\Delta_H + \epsilon_{\mathrm{plan}}.$$

Adding g_Γ proves Theorem 3; near-optimal policies give the same result when infima are not attained. Equal posterior interfaces realize the same policy class, proving Corollary 1.

Belief-state and information-value results. The factorization in Equation (15) is conditionally Markov given $\boldsymbol{\xi}_t$ and typed actions. Integrating future conditional laws against the exact posterior proves the standard belief-MDP reduction. For nonnegative VOI, the post-query decision class contains the constant rule that ignores the observation, so post-query risk cannot exceed prior risk.

Belief-coherence stability. For $0 \le f \le L$, $|\mathbb{E}_b f - \mathbb{E}_{b'} f| \le L\mathrm{TV}(b, b')$. Applying this inequality to a minimizer under the other belief and swapping the roles of b, b' proves the risk bound. A common observation kernel contracts total variation, so the same argument on the joint state–outcome law proves the post-query bound. The triangle inequality yields the VOI bound. Applying that bound to two queries and adding penalty changes proves query-order stability.

One-step acquisition and selective control. Uniform utility approximation and optimality of $\widehat{q}$ give $\mathcal{U}(b, q^\star) \le \widehat{\mathcal{U}}(b, q^\star) + \epsilon \le \widehat{\mathcal{U}}(b, \widehat{q}) + \epsilon \le \mathcal{U}(b, \widehat{q}) + 2\epsilon$. For false acceptance, on $A_\tau = 1$, $\widehat{U}_\delta \le \tau$, hence $\{\ell > \tau, A_\tau = 1\} \subseteq \{\ell > \widehat{U}_\delta\}$; calibration proves the theorem. Dividing by acceptance probability yields the conditional corollary.

Artifact availability. The package includes manuscript and figure sources, source-linked numerical inputs, exact analytic calculations, and validation receipts. Reproducible releases require configurations, checkpoints, predictions, unit counts, query traces, measured resources, and provenance.