TARGET-INDEPENDENT MICRO-INTERVENTIONS FOR PREDICTING TRAINING RESPONSE ACROSS LANGUAGE-MODEL FAMILIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Benchmark scores describe current capability but do not determine how a checkpoint responds to further training. We introduce L-STATE, which combines current scores with capability changes from four short, target-independent training interventions. Its measured responses support direct and structure-preserving operator readouts, with a cross-family prediction bound that accounts for coordinate heterogeneity. Relative to current scores alone, both pulse readouts reduce prediction mean squared error by 39.4% across three development families. Reductions reach 78.3% for the operator on held-out GLM and 78.4% for the direct readout on held-out Granite. Fixed eight-step probes also predict longer responses: on nine GLM checkpoints under repeated-example adaptation, a post-hoc evaluation without refitting or rescaling retains 63.3% and 47.9% error reductions at 16 and 32 target steps. In a separate GLM/Yi comparison across repeated- and fresh-example training, full L-STATE reduces prediction error by 36.9–38.8% and action-selection regret by 24.8–68.8% at a common 32-step endpoint. These results show that short, reusable interventions provide information beyond static capability, improving both response prediction and training choices.

1 INTRODUCTION

Two language-model checkpoints can have identical mathematics, code, science, reading, and language-modeling scores yet respond differently to further training. One may learn quickly, another may forget neighboring skills, and a third may barely change. Current scores record what a model does; controlled micro-training measures how its capabilities change after a short training episode. Here capability means negative clipped completion-token log loss on 32 held-out examples per coordinate, and response means its change under a short LoRA update. These continuous scores measure likelihood changes; task accuracy uses a different evaluation endpoint.

From the same checkpoint, we branch four independent, standardized, target-independent training episodes and measure each capability change. The four five-dimensional response vectors form a 20-dimensional pulse block. Appending the five current capability scores gives the full 25-dimensional L-STATE (Figure 1). The pulse block supports two cross-family readouts. The direct readout learns a multi-output map to the target-action response; the operator readout combines measured local response geometry with a source-estimated action coordinate. This common evaluation space compares families with different parameters, gradients, and hidden representations through the same measured responses.

To predict Qwen's response, a readout learns from Mistral and OLMo target responses and measures Qwen's capability and pulses. Qwen targets remain withheld until predictions are fixed. This leave-one-family-out (LOFO) design tests transfer through the common pulse measurement. Theorem 3.4 supports the operator under the conditions in Section 3; Eq. (10) separates target identification, source calibration, and cross-family coordinate bias.

In three-family LOFO development, both pulse readouts reduce the mean squared error (MSE) of predicted capability changes by 39.4% relative to using current scores alone. On GLM, direct reduces MSE by 71.8% and operator by 78.3%; the operator also improves balanced accuracy for predicting increase versus decrease by 38.8 percentage points. GLM's science-action inversion motivates action-

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

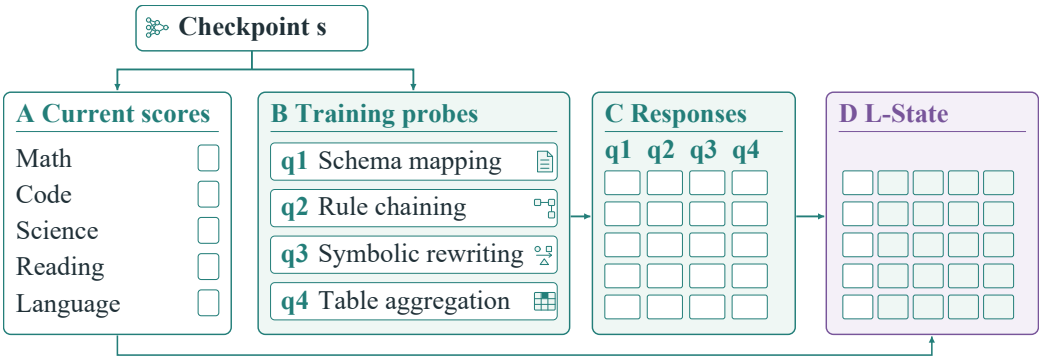

Figure 1: **Measuring L-State: current ability plus response to training.** **A**, $c(s)$ contains five held-out capability scores for mathematics, code, science, reading, and language modeling, measured by negative clipped token log loss; higher is better. **B**, four synthetic tasks probe how checkpoint s learns: q_1 maps a record to key-sorted JSON; q_2 follows a two-link rule; q_3 rewrites and joins symbols; q_4 adds two table lookups. Each branch uses 32 examples and eight LoRA steps, independently restarting from s with reset optimizer moments. Probe data are disjoint from candidate target tasks. **C**, reevaluating the same five capabilities gives after-minus-before score changes, forming the five-by-four response matrix $R_8(s)$. **D**, appending the five current scores to its 20 entries gives the 25-dimensional L-State. Grids show dimensions; the first column in D holds current scores and the other four hold probe responses.

wise selection from development folds before the Granite test. On Granite, direct and selector reduce prediction MSE by 78.4% and 77.7%; full L-STATE reduces the loss from choosing a suboptimal action by 25.5%.

A separate crossed evaluation asks whether this extra measurement improves training decisions beyond current capability. On GLM and Yi, each with nine checkpoints from three trajectories, full L-STATE reduces action-selection regret by 24.8–68.8% across repeated- and fresh-example settings at a common 32-step endpoint. We define regret against the best of the same four measured actions and evaluate choices using an independent final set. These comparisons measure the value of probe information over current scores; equal-time comparisons with adaptive screening price its acquisition cost (Section 5.7).

Our contributions are:

- We introduce the 25-dimensional L-STATE and its direct and operator readouts, with conditional response, sign, regret, and exact-action bounds that explicitly account for source- and target-family coordinate deviations.
- We validate cross-family prediction and action ranking in development and sealed GLM/Granite tests, and quantify coordinate heterogeneity through a five-family analysis that informs readout selection.
- We evaluate fixed eight-step probes at 16- and 32-step target horizons and quantify their decision value in GLM/Yi repeated- and fresh-example training. Source-selected nonlinear readouts and measured cost comparisons provide complementary controls.

2 RELATED WORK

Predicting adaptation. Static diagnostic probes can predict fine-tuning performance in studied NLP settings (Zhu et al., 2022); analyses of intermediate checkpoints reveal training regularities shared across scales (Xia et al., 2023); and observational scaling laws compress benchmark measurements into a low-dimensional capability space (Ruan et al., 2024). Transferability scores similarly ask which pretrained representation or checkpoint will adapt well (Nguyen et al., 2020; Munn & Wei, 2025). LogME scores target-label evidence from frozen pretrained features without fine-tuning (You et al., 2021). Resource-constrained model selection has also been formulated as extrapolating full fine-tuning performance from smaller-data runs through a rectified scaling law (Lin et al.,

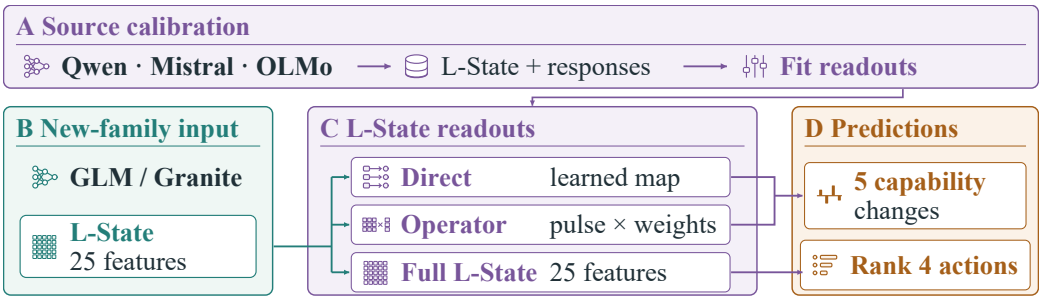

Figure 2: **From measured L-State to cross-family prediction and action selection.** **A**, Qwen, Mistral, and OLMo provide L-State measurements and observed responses to four candidate training tasks; readout fitting and setting selection use only these source families. **B**, a new GLM or Granite checkpoint supplies the measured L-State from Figure 1. **C**, direct and operator read the same 20 pulse responses: direct ridge predicts five capability changes, while the operator uses $R_8(s)\widehat{a}_u$, where $\widehat{a}_u \in \mathbb{R}^4$ is the source-fitted coordinate of candidate task u; full L-State uses all 25 features for action scoring. The action-wise selector chooses direct or operator using development-family folds. **D**, the outputs are predicted capability changes and utility-based rankings of four training actions: mathematics (GSM8K), code (MBPP), science (SciQ), and reading (BoolQ). Target responses are opened only after predictions and choices are fixed. One L-State measurement is reused across all four actions; detailed cost accounting appears in Appendix J.

2024). Model-level representation work constructs compact embeddings for correctness, routing, and benchmark-performance prediction (Zhuang et al., 2025), and training-free functional fingerprints compare heterogeneous models (Wu et al., 2026). These methods characterize current functional behavior; L-STATE derives its representation from controlled training interventions and predicts candidate-action response vectors. Closest in experimental goal, TUNEAHEAD combines dataset descriptors with a short standardized probe to forecast scalar final scores for runs based on Qwen (Luo et al., 2026). A concurrent preprint is closer in mechanism: it models learning through a parameter–optimizer receiving state and target-specific update geometry in nanoGPT, ResNet, and diffusion experiments (Wang, 2026). L-STATE differs in representation and validation: current capabilities and target-independent pulse responses live in a common evaluation space, predict a signed response vector over several candidate actions, and are evaluated by family-held-out development followed by chronologically sealed tests on GLM and Granite with downstream action regret.

From training data to behavior. Datamodels learn how training-set composition changes predictions (Ilyas et al., 2022); influence functions and optimizer-aware gradient methods trace or rank training examples by their target effect (Koh & Liang, 2017; Xia et al., 2024). Task arithmetic represents completed fine-tuning runs as directions in a shared parameter space (Ilharco et al., 2023). Recent work transports completed task vectors across heterogeneous-width models by aligning observed internal activations and their functional effect (Rinaldi et al., 2026). L-STATE takes a complementary route: it treats the checkpoint as the object being identified, measures target-independent pulse responses in evaluation space, and predicts future candidate-action responses. Our question is also conditional in the sense of Hewitt et al. (2021): do pulse responses provide usable information beyond current capability?

Intervention and domain transfer. Informative controlled inputs are central to active system identification (Wagenmaker & Jamieson, 2020). This analogy motivates standardized pulses, but our local response model states the training-domain assumptions explicitly. Following domain-generalization evaluation lessons (Gulrajani & Lopez-Paz, 2021), family identity defines the outer split and all selection occurs on source families. In the prospective tests, readout choices, response predictions, and utility-conditioned action choices are fixed before held-out-family target responses enter evaluation.

3 L-STATE AND ITS PULSE READOUTS

Method overview. Figure 2 shows how source calibration turns the measured L-State into target-response predictions.

3.1 WHY STATIC CAPABILITY IS INSUFFICIENT

Each model family m has a state space $\mathcal{X}_m$. We evaluate $x \in \mathcal{X}_m$ through a capability vector $c_m(x) \in \mathbb{R}^d$ in the same evaluation space for all families, with larger coordinates better. For a micro-training action u, strength h, and training randomness ξ, define the population response

$$r_m^h(x,u) = \frac{\mathbb{E}_\xi[c_m(U_{m,u}^h(x;\xi))] - c_m(x)}{h}. \tag{1}$$

Theorem 3.1 (Capability sufficiency and non-identifiability). *If $c_m(x) = c_{m'}(x') = c$, write $e = \|f(c,u) - r_m^h(x,u)\|_2$ and $e' = \|f(c,u) - r_{m'}^h(x',u)\|_2$. Every capability-only predictor satisfies*

$$\max\{e, e'\} \geq \tfrac{1}{2}\|r_m^h(x,u) - r_{m'}^h(x',u)\|_2. \tag{2}$$

Moreover, an exact capability-only response map exists on a domain if and only if, for every u, the response is constant on every capability level set.

The theorem characterizes exact level-set ambiguity. The empirical comparison instead tests prediction at observed checkpoints with continuous scores; local update directions and capability Jacobians can differ at similar score profiles. Controlled interventions measure this response variation.

3.2 LOCAL INTERVENTION GEOMETRY

The transfer bound requires five conditions: a named pulse gauge, smooth local dynamics, spanning pulse excitation, positive source coverage, and bounded measurement and misspecification errors. Whether target actions have the same coordinate across families is left as a hypothesis and enters the bounds quantitatively.

In the common named-pulse gauge, fix a reference family set and positive normalized weights $\{\pi_m\}_m$ and $\{\omega_u\}_u$. Write

$$a_{m,u} = a_0 + \alpha_u + \beta_m + \gamma_{m,u} = \bar{a}_u + \delta_{m,u}, \tag{3}$$

where $\bar{a}_u = a_0 + \alpha_u$ is pooled, β_m is the family deviation, and $\gamma_{m,u}$ is the family–action interaction. Weighted zero-sum centering (Eq. (15), Appendix B.2) gives $\bar{a}_u = \sum_m \pi_m a_{m,u}$ and $\delta_{m,u} = a_{m,u} - \bar{a}_u$. Exact sharing is the nested hypothesis

$$H_{\text{share}}: \quad \beta_m = 0 \text{ and } \gamma_{m,u} = 0 \quad \text{for all } m, u. \tag{4}$$

Under smooth local dynamics, the family-specific coordinate yields the response factorization $r_m^h(x,u) = B_m(x)a_{m,u} + d_m(x,u,h)$, where $B_m(x) = Jc_m(x)G_m(x)$. The residual separates update misspecification from finite-radius curvature; its explicit assumptions and bound are given in Theorem B.1 and Eqs. (12)–(14) (Appendix B.2). Training randomness enters through the expected displacement and its second moment. AdamW scaling, clipping, and LoRA parameterization enter the update-misspecification term ϵ_g; Appendix B.2 connects it to the pulse-mixture diagnostic.

3.3 IDENTIFYING THE CHECKPOINT RESPONSE OPERATOR

For a fixed family, let k target-independent pulses $q_1, \ldots, q_k$ have coordinate matrix $Q_m = [a_{m,q_1}, \ldots, a_{m,q_k}] \in \mathbb{R}^{r\times k}$. Their population and empirical response matrices satisfy

$$Y_m = B_m Q_m + D_m, \qquad \widehat{Y}_m = B_m Q_m + E_m, \qquad E_m = D_m + Z_m, \tag{5}$$

where D collects local residuals and Z collects training and evaluation noise.

Theorem 3.2 (Pulse identifiability). *For a fixed target coordinate $a_\star$, the response $Ba_\star$ is uniquely determined from BQ for every $B \in \mathbb{R}^{d\times r}$ if and only if $a_\star \in \mathrm{col}(Q)$. Hence every action in $\mathbb{R}^r$ is identifiable if and only if Q has full row rank. In that case, $\widehat{B} = \widehat{Y}Q^\dagger$ satisfies*

$$\|\widehat{B} - B\|_\nu \leq \frac{\|E\|_\nu}{\sigma_r(Q)}, \qquad \nu \in \{2, F\}. \tag{6}$$

The pulse-conditioning optimum is stated in Theorem B.2 (Appendix B.3).

Our estimator uses the same four named standard pulses as coordinate anchors, giving $Q_m = I_4$ separately in every family. This gauge convention aligns the meanings of the four axes. Cross-family equality of target-action coefficients remains the separate hypothesis H_{share}. The convention also removes an explicit $Q_m^\dagger$ factor from operator recovery. With an independently fixed action metric and energy budget, Theorem B.2 selects tight-frame interventions. With common pulse strength h, the stored differences $R = h\widehat{Y}$ change only the global scale. The empirical state is

$$z_m(x) = \left[\widehat{c}_m(x); \operatorname{vec} R_m(x)\right] \in \mathbb{R}^{25}. \tag{7}$$

Every pulse branches from an independent copy of the checkpoint.

3.4 Estimating a Pooled Action under Family Heterogeneity

For source checkpoints $\mathcal{S} = \{s_1, \ldots, s_n\}$, let $\mathcal{B}_S$ stack their operators and let $v_{S,u}^{\mathrm{het}}$ stack their family-specific response deviations. The resulting source model is $y_S(u) = \mathcal{B}_S \bar{a}_u + v_{S,u}^{\mathrm{het}} + e_S$; Eq. (19) gives the full construction. The source-coverage radius is $\underline{\sigma}_S = \sigma_r(\mathcal{B}_S)$, while $H_{S,u} = \|v_{S,u}^{\mathrm{het}}\|_2$ measures the linear response-scale cost of imposing a pooled coordinate on heterogeneous source families.

Theorem 3.3 (Approximate pooled-coordinate recovery under bounded heterogeneity)**.** *Assume $\underline{\sigma}_S > 0$, $\|e_S\|_2 \le \epsilon_S$, $\|\bar{a}_u\|_2 \le A$, and $\|v_{S,u}^{\mathrm{het}}\|_2 \le H_{S,u}$. With exact source operators, $\widehat{a}_u = \mathcal{B}_S^\dagger y_S(u)$ obeys $\|\widehat{a}_u - \bar{a}_u\|_2 \le (H_{S,u} + \epsilon_S)/\underline{\sigma}_S$. If $\|\widehat{\mathcal{B}}_S - \mathcal{B}_S\|_2 \le \tau_S < \underline{\sigma}_S$, then*

$$\widehat{a}_u = \widehat{\mathcal{B}}_S^\dagger y_S(u), \qquad \|\widehat{a}_u - \bar{a}_u\|_2 \le \frac{\tau_S A + H_{S,u} + \epsilon_S}{\underline{\sigma}_S - \tau_S}. \tag{8}$$

Under H_{share}, $H_{S,u} = 0$ and the original shared-coordinate bound is recovered as a special case. Here $H_{S,u}$ measures deviations after projection through the observed source operators; family deviations in their null spaces can coexist with $H_{S,u} = 0$.

Additional sources tighten the complete bound when their coverage gain outpaces operator, response, and coordinate-heterogeneity errors. Persistent heterogeneity can leave a nonzero pooling-error floor even as coverage grows (Appendix B.4).

3.5 Prospective Transfer to an Unseen Family

Let $\star$ denote a previously unseen family. Its target-independent pulses give $\widehat{B}_\star$, while source-family target responses give $\widehat{a}_u$, an estimate of the pooled coordinate $\bar{a}_u$. The prediction is $\widehat{r}_\star(u) = \widehat{B}_\star \widehat{a}_u$.

Theorem 3.4 (End-to-end cross-family transfer)**.** *Let $Q_\star$ be expressed in the same fixed named-pulse gauge, have full row rank, and let $\widehat{B}_\star = \widehat{Y}_\star Q_\star^\dagger$. Assume $\|\bar{a}_u\|_2 \le A$, $a_{\star,u} = \bar{a}_u + \delta_{\star,u}$ with $\|\delta_{\star,u}\|_2 \le D_{\star,u}$, $\|r_\star^h(u) - B_\star a_{\star,u}\|_2 \le \epsilon_{\star,u}$, target pulse perturbation $\|E_\star\|_2 \le \eta_\star$, and the source conditions of Theorem 3.3. Then*

$$\|\widehat{r}_\star(u) - r_\star^h(u)\|_2 \le \Delta_{\star,u}, \tag{9}$$

where

$$\begin{aligned} \Delta_{\star,u} = {} & \frac{\eta_\star A}{\sigma_r(Q_\star)} + \epsilon_{\star,u} \\ & + \|B_\star\|_2 D_{\star,u} \\ & + \left(\|B_\star\|_2 + \frac{\eta_\star}{\sigma_r(Q_\star)}\right) \frac{\tau_S A + H_{S,u} + \epsilon_S}{\underline{\sigma}_S - \tau_S}. \end{aligned} \tag{10}$$

Equation (10) separates target identification, target locality, unseen-family coordinate bias, and source heterogeneity. The proof bounds pulse inversion and source-coordinate estimation, then propagates both errors through $\widehat{B}_\star \widehat{a}_u$. Appendix B gives the full proofs, finite-repeat bounds, and the ridge and realized-response versions.

Corollary 3.5 (Operator direction and action recovery)**.** *If* $|r^h_{\star,j}(u)| > \Delta_{\star,u}$*, then* $\operatorname{sign} \widehat{r}_{\star,j}(u) = \operatorname{sign} r^h_{\star,j}(u)$*. For a finite candidate set with* $J(u) = hv^\top r^h_\star(u) - \operatorname{cost}(u)$*, let* $u^\star \in \arg\max_u J(u)$ *and* $\widehat{u} \in \arg\max_u \widehat{J}(u)$*, where* $\widehat{J}(u) = hv^\top \widehat{r}_\star(u) - \operatorname{cost}(u)$*. Simultaneous radii imply*

$$J(u^\star) - J(\widehat{u}) \le h\|v\|_2 \big(\Delta_{\star,u^\star} + \Delta_{\star,\widehat{u}}\big). \tag{11}$$

If the true best-versus-runner-up utility margin exceeds $h\|v\|_2(\Delta_{\star,u^\star} + \max_{u \neq u^\star} \Delta_{\star,u})$*, then* $\widehat{u} = u^\star$*. Balanced-accuracy, normalized-regret, and realized-response versions appear in Appendix B.*

3.6 Reading Training Responses from L-State

Direct pulse readout. For each target action, source-only multi-output ridge maps the 20-dimensional pulse block $p(s) = \operatorname{vec} R(s)$ to the five-dimensional response, allowing unconstrained linear combinations of the measured pulses.

Operator readout. Source ridge fitting estimates $\widehat{a}_u$ and predicts $\widehat{y}_{\mathrm{op}}(s,u) = R(s)\widehat{a}_u$, preserving the zero point and local operator geometry. For this ridge estimator, replace the final fraction in Eq. (10) by $b_\lambda A + \rho_\lambda(\tau_S A + H_{S,u} + \epsilon_S)$ from Eq. (20), including shrinkage bias. The displayed theorem uses unpenalized source fitting.

Full L-State readout. An action-specific multi-output ridge map uses $z(s) = [c(s); p(s)]$. Source-trajectory validation selects regularization; Appendix C specifies the transformations. Corollary 3.5 instantiates operator error bounds; full-state action gains are evaluated empirically. The radius is conditional on unseen-family residual and heterogeneity bounds (Appendix B).

Shared-coordinate hypothesis tests. After all labels open, one coordinate per family–action cell tests Eq. (3). Equal family weights define the retrospective mean $\bar{a}^{(5)}_u$, distinct from the prospective mean. Difference, retrospective equivalence, and predictive tests use unpenalized least squares; only stability summaries use ridge (Section 5.5).

Action-wise pulse selector. The first sealed family motivates action-wise selection using only the three held-out development-family folds. Direct is selected after a strict RMSE win in at least two folds; ties retain operator. This response rule gives operator for mathematics/reading and direct for code/science. The mapping is frozen before Granite target training. We use this selector to predict responses, the operator to predict direction, and full state to score actions.

4 Experimental Setup

4.1 Development Families and Sealed Families

Development uses Qwen (Qwen Team, 2025), Mistral (Jiang et al., 2023; Mistral AI, 2024), and OLMo (Walsh et al., 2025); held-out tests use GLM (Zhipu AI, 2025) and then Granite (IBM Granite Team, 2024). These five 7–9B instruction-tuned families provide 135 checkpoint states. Each state-generation trajectory runs for 64 optimizer steps. Appendix C gives checkpoint spacing, seeds, and family-level splits.

4.2 Data Separation and Training Settings

Data have disjoint roles. Four deterministic synthetic tasks generate state trajectories. Four different synthetic tasks provide the standard pulses: schema mapping, two-step rule chaining, symbolic rewriting, and table aggregation; this pulse set is synthetic and disjoint from the target benchmark tasks. Semantic target actions train on 32 examples from GSM8K (Cobbe et al., 2021), MBPP (Austin et al., 2021), SciQ (Welbl et al., 2017), or BoolQ (Clark et al., 2019). Capabilities are negative clipped token NLL on disjoint held-out examples from those four tasks plus WikiText-2 (Merity et al., 2017). Normalized texts are disjoint across state generation, pulse training, target training, and capability evaluation.

All primary pulse and target branches freeze base weights and train rank-4 LoRA on attention query and value projections for eight AdamW steps, consuming 32 examples. The target-early reference reads a target-specific intermediate after two steps and eight examples, while locality calibration deliberately includes four- and sixteen-step variants. Each branch starts with reset optimizer moments and restores the identical checkpoint. These short LoRA updates measure the checkpoint's immediate training response.

4.3 Metrics and Uncertainty

Response error measures how far predicted capability changes are from their observed values. RMSE uses source-only means and population standard deviations within each outer fold, action, and coordinate. We report relative MSE reduction as $100(1 - \mathrm{RMSE}_m^2/\mathrm{RMSE}_{\mathrm{cap}}^2)\%$ against the capability-only reference: a 50% reduction means half as much squared prediction error. Sign balanced accuracy uses coordinate-specific dead zones from three-repeat pulse reliability; GLM has 184 determinate components, and Granite keeps the same source-frozen thresholds. Appendix C.6 gives their formula and source population. Continuous RMSE retains all components. We also report median response cosine. Six utilities (five one-hot capabilities and one balanced vector) rank four semantic actions, measured by top-1 accuracy (choosing the best action) and realized normalized regret (utility lost relative to that action) under one preregistered target realization. Regret reductions are relative percentages; changes in accuracy are percentage-point differences. Family splits, baseline definitions, and prediction-freezing procedures appear in Appendix C.1.

Development uses 10,000 family-then-trajectory bootstrap resamples for full-state metrics and 1,000 aligned-pulse permutations with source refitting. Each sealed core analysis uses 10,000 whole-trajectory resamples over three trajectories; the frozen selector has pooled and action-macro point estimates. Four extra GLM science responses per state check repeat stability; a separate nine-state study checks four/eight/sixteen-step exposure. Independent reruns reproduce all three stages.

5 Results

The proposed predictors all build on L-State. Direct and operator are two ways to read its 20 measured pulse responses; full L-State also uses the five current capability scores to rank actions. These measurements reduce response-prediction MSE by 39.4% with either pulse readout in development. The reduction reaches 78.3% for the operator on GLM and 78.4% for the direct readout on Granite, relative to current scores alone (Table 1).

Among the L-State readouts, direct has the lowest response RMSE in development and Granite, and operator has the lowest on GLM. The matched-state reference has the lowest action regret in all three stages, using source responses matched by seed, state index, and action.

5.1 L-State Readouts in Three-Family Development

Both readouts of the L-State pulse block reduce response MSE by 39.4%; direct has the lowest RMSE and operator the highest sign BA. Full L-State reduces the utility lost through action selection by 47.5%, with normalized regret falling from 0.472 to 0.248. The best way to read L-State therefore depends on whether the goal is response prediction or action selection. Aligned-pulse refits exceed the registered per-fold null 95th percentile on Qwen and Mistral. Appendix D gives fold-level results and ablations.

5.2 Sealed GLM Test of the Operator Readout

For GLM, we fix 540 response predictions and 810 action choices from the L-State readouts before target labels open. Direct and operator MSE gains are 71.8% and 78.3% relative to capability alone. The operator raises direction balanced accuracy from 36.6% to 75.4%, a gain of 38.8 percentage points, and has the highest core sign BA (Table 1). Full-state regret improves by 26.7%, from 0.592 to 0.434. Whole-trajectory 95% percentile intervals are $[78.1\%, 78.5\%]$ for operator MSE gain, $[0.705, 0.860]$ for its BA, and $[25.0\%, 29.9\%]$ for full-state regret reduction, conditional on the three observed trajectories. Appendix E gives complete resampling and action-wise results.

Table 1: **Core response, direction, and action comparisons.** Development: 81 states, three held-out-family folds; each sealed family: 27 states, three trajectories. Compare source-standardized RMSE within each stage. BA is sign balanced accuracy; regret uses six utilities. Bold: best core point estimate. Dashes: outside this core comparison; the frozen Granite selector's auxiliary sign metrics appear in Appendix F.

Readout	Development			Sealed GLM			Sealed Granite		
	RMSE ↓	BA ↑	Regret ↓	RMSE ↓	BA ↑	Regret ↓	RMSE ↓	BA ↑	Regret ↓
Source mean	7.895	0.479	0.268	–	–	–	–	–	–
Capability	9.987	0.602	0.472	1.339	0.366	0.592	1.172	0.654	0.291
Direct pulse	**7.772**	0.531	0.299	0.712	0.571	0.451	**0.544**	0.535	0.215
Scalar statistics	8.738	0.547	0.365	–	–	–	–	–	–
Matched state	7.913	0.491	**0.232**	0.939	0.569	**0.410**	0.692	0.535	**0.212**
Operator	7.777	**0.675**	0.303	**0.624**	**0.754**	0.513	0.634	0.657	0.253
Full L-State	8.293	0.589	0.248	0.841	0.359	0.434	0.680	**0.664**	0.217
Target early (2 steps)	8.067	0.671	0.283	–	–	–	–	–	–
Directional blend	7.876	0.643	0.292	–	–	–	–	–	–
Action-wise selector	–	–	–	–	–	–	0.554	–	–

5.3 Action-Wise Readout Selection

On 28 determinate GLM science-action coordinates, operator sign BA is zero and direct BA is 1.0; four further realizations retain the target sign in all 27 states. Development folds select operator for mathematics/reading and direct for code/science before Granite labels open. Appendix F gives the repeat and coordinate diagnostics.

5.4 Sealed Granite Test of Action-Wise Readout Selection

On Granite, the direct and action-wise selected L-State readouts reduce prediction MSE by 78.4% and 77.7% relative to capability alone. Their RMSEs are 0.544 and 0.554, compared with 0.634 for the global operator and 1.172 for capability (Table 1). Operator sign BA is 0.657 versus capability's 0.654; its gain interval $[-0.0615, 0.0823]$ includes zero. Full-state regret falls by 25.5% ($[15.7\%, 40.1\%]$), and best-action selection accuracy rises by 10.5 percentage points, from 53.1% to 63.6%. Appendix H gives all metrics and trajectory-conditional intervals.

5.5 Testing the Operator Readout's Action Coordinate

The five-family analysis rejects exact sharing for code, science, and reading at level 0.05; mathematics remains unresolved. Family–action coordinates reduce held-trajectory MSE by 17.3% overall and 48.0% for science, supporting explicit coordinate-bias terms (Table 20; Appendix G).

5.6 Predicting Longer-Horizon Training Responses from Short Probes

Fixed eight-step probes predict longer total responses on nine GLM checkpoints under repeated-example LoRA adaptation (Table 24). On the original evaluation set, later used for selection, reusing the original predictions unchanged gives operator MSE reductions of 63.3% at 16 target steps and 47.9% at 32, without horizon-specific refitting or rescaling. At 32 steps, operator sign BA is 76.2% versus capability's 47.5%, a gain of 28.6 percentage points in predicting increase versus decrease. On the independent final set, the 32-step operator BA is 65.4% versus 42.9% for capability (Table 39). Full L-State lowers action-selection loss by 30.5% at 16 steps and 36.6% at 32, relative to capability alone.

With pulses also extended to 32/64/128 steps, matched-horizon operator MSE reductions are 53.9%, 54.1%, and 54.2% with fixed source fits. Compact repeated-data dynamics govern this comparison;

Table 2: **Incremental value over current capability at a common 32-step target endpoint.** Full four-probe L-STATE versus the original capability-only baseline, using independent final outcomes. ΔU is the paired mean balanced-utility difference, followed by a descriptive trajectory-level interval. MSE and regret columns give relative reductions; W/T/L counts better/equal/worse state-level utility. Each row has nine states from three trajectories. The best measured action defines retrospective regret and does not enter selection.

Family	Setting	ΔU [interval]	MSE ↓	Regret ↓	W/T/L
GLM	Repeated	$+0.1592$ $[0.1458, 0.1771]$	38.8%	40.2%	3/6/0
GLM	Fresh	$+0.1105$ $[0.0851, 0.1383]$	38.3%	24.8%	3/6/0
Yi	Repeated	$+0.7188$ $[0.6265, 0.8729]$	36.9%	68.8%	3/6/0
Yi	Fresh	$+0.7714$ $[0.6614, 0.9273]$	38.5%	46.3%	3/6/0

the fresh-example study separately checks continued data exposure. Appendices I–J give results and costs.

5.7 VALUE OF PROBE INFORMATION BEYOND CURRENT CAPABILITY

We compare four probes with the original capability-only baseline at a common 32-step target endpoint. GLM and Yi each contribute nine states from three trajectories; repeated- and fresh-example continuations share their first eight steps. The readout and action rule remain source-frozen. Independent final measurements show 36.9–38.8% lower response MSE and 24.8–68.8% lower action-selection regret (Table 2). Regret is $\max_{a \in \mathcal{A}} U_{32}(s, a) - U_{32}(s, \widehat{a})$ over the same four actions; retrospective best actions enter scoring only. Utility weights $(1, 1, 1, 1, 0.5)$ combine five negative clipped-token-NLL coordinates.

Each cell has three improved and six unchanged decisions. The corrections select code instead of mathematics at GLM's initial states and instead of science at Yi's later observed states. The same measurement can serve multiple queries at an unchanged checkpoint. Fixed-endpoint information gains and equal-time deployment comparisons answer different questions: Table 29 reports the delivered utilities and acquisition costs alongside target-specific early screening and successive halving. Appendix K.1 reports hardware, timing, and budget-dependent cost–benefit comparisons. Source-selected linear, quadratic, and radial-basis controls also improve mean balanced utility when given probes in all four cells under both matched selection protocols (Appendix K.6). These finite comparisons complement Theorem 3.1; they do not identify the optimum over all capability-only predictors.

6 CONCLUSION

Target-independent micro-interventions measure a transferable training-response state. Direct and operator readouts improve cross-family prediction, while theory and five-family analysis quantify operator-coordinate heterogeneity; full L-STATE also improves action ranking. On nine GLM checkpoints under repeated-example adaptation, fixed eight-step probes predict 16- and 32-step responses without refitting or rescaling. At 32 steps, the operator retains a 47.9% MSE reduction over capability: the short measurements remain predictive four times farther into training than their per-probe duration.

The GLM/Yi comparison further links this information to training decisions. Across repeated- and fresh-example settings at a common 32-step endpoint, full L-STATE reduces response MSE by 36.9–38.8% and action-selection regret by 24.8–68.8% relative to the capability-only baseline. These gains establish the value of probe information for the evaluated choices. The same measurement can serve multiple queries at an unchanged checkpoint; the measured cost comparisons quantify its acquisition expense.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript writing and language polishing, including improving clarity, grammar, and phrasing. The authors take responsibility for the final content of the paper, including its claims and conclusions.

REPRODUCIBILITY STATEMENT

Sections 4–5 describe the data separation, training settings, prediction timing, metrics, and independent reproduction. Appendices A–K provide complete proofs, experimental details, development and sealed-family tables, repeat and duration studies, incremental-cost accounting, crossed-setting comparisons, and reproducibility details.

REFERENCES

Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. Program synthesis with large language models. arXiv preprint arXiv:2108.07732, 2021. URL https://arxiv.org/abs/2108.07732.

Christopher Clark, Kenton Lee, Ming-Wei Chang, Tom Kwiatkowski, Michael Collins, and Kristina Toutanova. BoolQ: Exploring the surprising difficulty of natural yes/no questions. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 2924–2936. Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-1300. URL https://aclanthology.org/N19-1300/.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. arXiv preprint arXiv:2110.14168, 2021. URL https://arxiv.org/abs/2110.14168.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=lQdXeXDoWtI.

John Hewitt, Kawin Ethayarajh, Percy Liang, and Christopher D. Manning. Conditional probing: Measuring usable information beyond a baseline. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 1626–1639. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.emnlp-main.122. URL https://aclanthology.org/2021.emnlp-main.122/.

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

IBM Granite Team. Granite-3.1-8B-Instruct model card. Hugging Face model card, 2024. URL https://huggingface.co/ibm-granite/granite-3.1-8b-instruct.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=6t0Kwf8-jrj.

Andrew Ilyas, Sung Min Park, Logan Engstrom, Guillaume Leclerc, and Aleksander Madry. Datamodels: Understanding predictions with data and data with predictions. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9525–9587. PMLR, 2022. URL https://proceedings.mlr.press/v162/ilyas22a.html.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7B. arXiv preprint arXiv:2310.06825, 2023. URL https://arxiv.org/abs/2310.06825.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1885–1894. PMLR, 2017. URL https://proceedings.mlr.press/v70/koh17a.html.

Haowei Lin, Baizhou Huang, Haotian Ye, Qinyu Chen, Zihao Wang, Sujian Li, Jianzhu Ma, Xiaojun Wan, James Zou, and Yitao Liang. Selecting large language model to fine-tune via rectified scaling law. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30080–30107. PMLR, 2024. URL https://proceedings.mlr.press/v235/lin24j.html.

Yuxiang Luo, Haonan Long, Chen Wang, Qiqi Duan, Xiaotian Lin, Yanwei Xu, Yuyu Luo, Weikai Yang, and Nan Tang. TuneAhead: Predicting fine-tuning performance before training begins. In *International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/64847.

Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. Pointer sentinel mixture models. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=Byj72udxe.

Mistral AI. Mistral-7B-Instruct-v0.3 model card. Hugging Face model card, 2024. URL https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3.

Michael Munn and Susan Wei. A Bayesian model selection criterion for selecting pretraining checkpoints. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 45256–45271. PMLR, 2025. URL https://proceedings.mlr.press/v267/munn25a.html.

Cuong Nguyen, Tal Hassner, Matthias Seeger, and Cedric Archambeau. LEEP: A new measure to evaluate transferability of learned representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7294–7305. PMLR, 2020. URL https://proceedings.mlr.press/v119/nguyen20b.html.

Qwen Team. Qwen2.5 technical report. arXiv preprint arXiv:2412.15115v2, 2025. URL https://arxiv.org/abs/2412.15115v2.

Filippo Rinaldi, Aniello Panariello, Giacomo Salici, Angelo Porrello, and Simone Calderara. Transporting task vectors across different architectures without training. In *International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/62524.

Yangjun Ruan, Chris J. Maddison, and Tatsunori Hashimoto. Observational scaling laws and the predictability of language model performance. In *Advances in Neural Information Processing Systems*, volume 37, pp. 15841–15892, 2024. doi: 10.52202/079017-0506. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/1cded4f97cf5f01a284c574110b7e3b9-Abstract-Conference.html.

Andrew Wagenmaker and Kevin Jamieson. Active learning for identification of linear dynamical systems. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 3487–3582. PMLR, 2020. URL https://proceedings.mlr.press/v125/wagenmaker20a.html.

Evan Pete Walsh, Luca Soldaini, Dirk Groeneveld, Kyle Lo, Shane Arora, Akshita Bhagia, Yuling Gu, Shengyi Huang, Matt Jordan, et al. 2 OLMo 2 furious (COLM's version). In *Second Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=2ezugTT9kU.

Mian Wang. Training, learning and inference: Unified dynamics of neural systems. arXiv preprint arXiv:2608.20965, 2026. URL https://arxiv.org/abs/2608.20965.

Johannes Welbl, Nelson F. Liu, and Matt Gardner. Crowdsourcing multiple choice science questions. In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pp. 94–106. Association for Computational Linguistics, 2017. doi: 10.18653/v1/W17-4413. URL https://aclanthology.org/W17-4413/.

Zhaomin Wu, Haodong Zhao, Ziyang Wang, Jizhou Guo, Qian Wang, and Bingsheng He. LLM DNA: Tracing model evolution via functional representations. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=UIxHaAqFqQ.

Mengzhou Xia, Mikel Artetxe, Chunting Zhou, Xi Victoria Lin, Ramakanth Pasunuru, Danqi Chen, Luke Zettlemoyer, and Veselin Stoyanov. Training trajectories of language models across scales. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 13711–13738. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.acl-long.767. URL https://aclanthology.org/2023.acl-long.767/.

Mengzhou Xia, Sadhika Malladi, Suchin Gururangan, Sanjeev Arora, and Danqi Chen. LESS: Selecting influential data for targeted instruction tuning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 54104–54132. PMLR, 2024. URL https://proceedings.mlr.press/v235/xia24c.html.

Kaichao You, Yong Liu, Jianmin Wang, and Mingsheng Long. Logme: Practical assessment of pre-trained models for transfer learning. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 12133–12143. PMLR, 2021. URL https://proceedings.mlr.press/v139/you21b.html.

Zhipu AI. GLM-4-9B-0414 model card. Hugging Face model card, 2025. URL https://huggingface.co/zai-org/GLM-4-9B-0414.

Zining Zhu, Soroosh Shahtalebi, and Frank Rudzicz. Predicting fine-tuning performance with probing. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 11534–11547. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.emnlp-main.793. URL https://aclanthology.org/2022.emnlp-main.793/.

Richard Zhuang, Tianhao Wu, Zhaojin Wen, Andrew Li, Jiantao Jiao, and Kannan Ramchandran. EmbedLLM: Learning compact representations of large language models. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/bf5e4b85d203481d6e37bd32d9600162-Abstract-Conference.html.

A Sealed-Family Summary Tables

B Core Proofs

B.1 Capability Level Sets

Proof of Theorem 3.1. Let $r = r^h_m(x,u)$, $r' = r^h_{m'}(x',u)$, and $y = f(c,u)$. The triangle inequality gives $\|r - r'\|_2 \le \|r - y\|_2 + \|y - r'\|_2$, so at least one of the two errors is at least $\|r - r'\|_2/2$.

For the characterization, if $r^h_m(x,u) = f(c_m(x),u)$, equal capabilities imply equal responses. Conversely, if responses are constant on every capability level set, define $f(c,u)$ as the common response of any checkpoint in that level set. Constancy makes this definition independent of the chosen representative. □

B.2 Factorization from Smooth Dynamics

Work in a local Euclidean chart around x and write the random training displacement as $\Delta^h_{m,u}(x,\xi)$. Let the capability Jacobian be L_c-Lipschitz along every segment traversed by this displacement. Suppose the family-specific coordinate $a_{m,u} \in \mathbb{R}^r$ and a checkpoint-specific update map $G_m(x)$ obey

$$\begin{aligned} \left\| \frac{\mathbb{E}\Delta^h_{m,u}}{h} - G_m(x)a_{m,u} \right\|_2 &\le \epsilon_g, \\ \mathbb{E}\|\Delta^h_{m,u}\|_2^2 &\le h^2 V^2 \|a_{m,u}\|_2^2. \end{aligned} \tag{12}$$

Theorem B.1 (Smooth dynamics induce the response factorization)**.** *Under Eq.* (12)*, the response in Eq.* (1) *has the form*

$$\begin{aligned} r^h_m(x,u) &= B_m(x)a_{m,u} + d_m(x,u,h), \\ B_m(x) &= Jc_m(x)G_m(x), \end{aligned} \tag{13}$$

with

$$\|d_m(x,u,h)\|_2 \le \|Jc_m(x)\|_2 \epsilon_g + \tfrac{1}{2} L_c h V^2 \|a_{m,u}\|_2^2. \tag{14}$$

The first term measures departure of the expected update from the family-specific coordinate factorization; the second is the finite-radius curvature cost. Training randomness enters through the expected displacement and its second moment, while repeat averaging controls the empirical estimation noise below. Cross-family sharing is governed separately by Eq. (4).

The decomposition in Eq. (3) is identified by

$$\sum_u \omega_u \alpha_u = 0, \quad \sum_m \pi_m \beta_m = 0, \quad \sum_m \pi_m \gamma_{m,u} = 0 \ \forall u, \quad \sum_u \omega_u \gamma_{m,u} = 0 \ \forall m. \tag{15}$$

These conditions imply $\bar{a}_u = \sum_m \pi_m a_{m,u}$ and $\delta_{m,u} = a_{m,u} - \bar{a}_u$.

Proof of Theorem B.1. Suppress (m,x,u,h) and write $\Delta = \Delta^h_{m,u}(x,\xi)$. The integral Taylor formula gives

$$c_m(x+\Delta) - c_m(x) = Jc_m(x)\Delta + \rho(\Delta), \tag{16}$$

where

$$\begin{aligned} \|\rho(\Delta)\|_2 &\le \int_0^1 \|Jc_m(x+t\Delta) - Jc_m(x)\|_2 \, \|\Delta\|_2 \, dt \\ &\le \tfrac{1}{2} L_c \|\Delta\|_2^2. \end{aligned}$$

Taking expectations, dividing by h, and inserting $\mathbb{E}\Delta/h = G_m(x)a_{m,u} + e_g$ yields

$$r^h_m(x,u) = Jc_m(x)G_m(x)a_{m,u} + Jc_m(x)e_g + \mathbb{E}\rho(\Delta)/h.$$

The assumed bounds on e_g and the second moment of Δ give Eq. (14). □

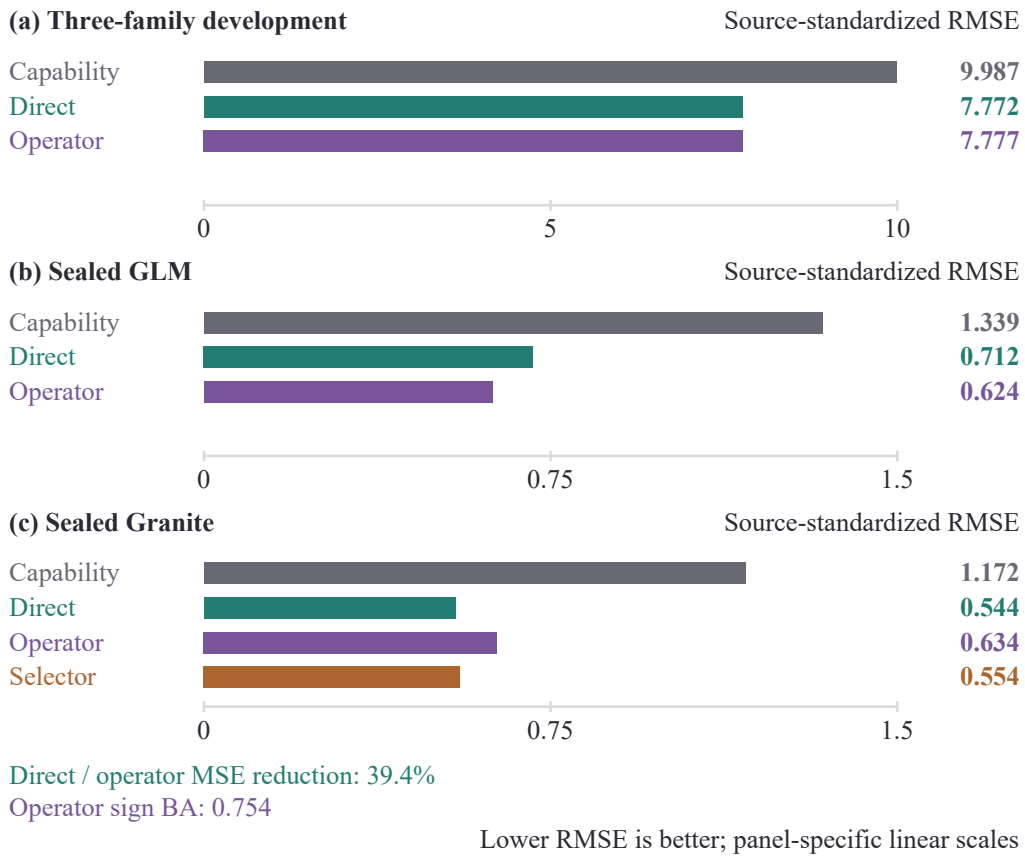

Figure 3: **Response prediction across model families.** Bars show source-standardized RMSE point estimates on separate linear scales. Both pulse readouts reduce development MSE by 39.4% versus capability. The Granite selector is fitted on development families; uncertainty analyses appear in Appendices E and H.

For the implemented optimizer, coordinate-wise moment normalization, global gradient clipping, and the LoRA parameterization can all contribute to ϵ_g in Eq. (12). Its size is an assumption distinct from the $O(h)$ curvature term. The measured median pulse-mixture relative residual of 0.342 (Appendix I) summarizes discrepancy in response space; it combines these effects with curvature and measurement noise and does not separately estimate ϵ_g.

For deterministic gradient flow with primitive losses $L_{a_m} = \sum_{j=1}^{r} a_{m,j} L_j$, the update map at x has columns $-\nabla L_j(x)$. The theorem then recovers $B_{:,j} = -Jc(x)\nabla L_j(x)$, with an $O(h\|a_m\|_2^2)$ remainder whenever the induced capability velocity is locally Lipschitz.

B.3 Pulse Identifiability and Design

Theorem B.2 (Minimax pulse conditioning)**.** *For full-row-rank Q, the exact worst-case spectral amplification is*

$$\sup_{\|E\|_2 \le 1} \|EQ^\dagger\|_2 = \frac{1}{\sigma_r(Q)}. \tag{17}$$

Under $\|Q\|_F^2 \le E_0$ and $k \ge r$, its minimum is $\sqrt{r/E_0}$, attained exactly by tight frames satisfying $QQ^\top = (E_0/r)I_r$.

Proof of Theorem 3.2. If $a_\star = Q\lambda$, then $Ba_\star = (BQ)\lambda$, so the pulse observations determine the target response. If $a_\star \notin \mathrm{col}(Q)$, choose $z \in \mathrm{null}(Q^\top)$ with $z^\top a_\star \neq 0$ and a nonzero $w \in \mathbb{R}^d$. The

Table 3: **First sealed-family evaluation.** Predictions and action choices are frozen before GLM target-action labels are opened. The direct readout uses the L-State pulse block as features; the operator readout is the frozen structured response/direction predictor, and the full L-State readout is the frozen action predictor. Sign BA is coordinate-macro balanced accuracy over source-dead-zone-determinate components. The direct and matched-state methods are compared on the same response and action-selection metrics.

Signed continuous response and direction

Method	RMSE ↓	MSE gain ↑	Sign BA ↑
Capability	1.339	0.000	0.366
Direct pulse	0.712	0.718	0.571
Operator†	**0.624**	**0.783**	**0.754**

Action selection

Method	Regret ↓	Reduction ↑	Top-1 ↑
Capability	0.592	0.000	0.340
Matched state	**0.410**	**0.307**	**0.432**
Full L-State†	0.434	0.267	0.395

†Readout fixed before testing. MSE gain and regret reduction use capability as the reference.

matrix $H = wz^\top$ obeys $HQ = 0$ but $Ha_\star \neq 0$; hence B and $B + H$ have identical pulse responses and different target responses. Applying this statement to every $a_\star \in \mathbb{R}^r$ gives the full-row-rank characterization.

When Q has full row rank, $QQ^\dagger = I_r$, and therefore

$$\widehat{B} - B = (BQ + E)Q^\dagger - B = EQ^\dagger.$$

Submultiplicativity and $\|Q^\dagger\|_2 = 1/\sigma_r(Q)$ give both norm bounds in Eq. (6). □

Proof of Theorem B.2. The upper bound $\|EQ^\dagger\|_2 \leq \|E\|_2\|Q^\dagger\|_2$ is attained by choosing a rank-one E aligned with a leading left singular vector of $Q^\dagger$. Thus the exact amplification is $\|Q^\dagger\|_2 = 1/\sigma_r(Q)$.

Let $\sigma_1 \geq \cdots \geq \sigma_r > 0$ be the singular values of Q. Then

$$\sigma_r(Q)^2 \leq \frac{1}{r}\sum_{j=1}^{r}\sigma_j(Q)^2 = \frac{\|Q\|_F^2}{r} \leq \frac{E_0}{r}.$$

Equality holds exactly when the energy budget is tight and all r singular values are equal, equivalently $QQ^\top = (E_0/r)I_r$. Taking reciprocals proves the minimax statement. □

If each of k pulses is independently repeated n times, every coordinate of the repeat-mean noise is mean-zero sub-Gaussian with variance proxy σ^2/n, and each pulse residual has norm at most ϵ_p, then with probability at least $1 - \delta$,

$$\|E\|_2 \leq \sqrt{k}\epsilon_p + \sigma\sqrt{\frac{2dk\log(2dk/\delta)}{n}}. \tag{18}$$

This follows from a coordinatewise tail bound, a union bound over dk entries, and $\|Z\|_2 \leq \|Z\|_F$. Combining Eqs. (6) and (18) supplies a finite-repeat choice of the target-pulse term $\eta_\star$ in Theorem 3.4; the source target-response, target-action residual, and realized-response terms retain their displayed radii.

Table 4: **Second sealed-family evaluation on Granite.** The action-wise selector is fitted only from the three development families and frozen before Granite target branches; the operator readout is registered for direction and the full L-STATE readout for action ranking. All rows are frozen before target-action labels are opened. Full five-method metrics appear in Appendix H.

Continuous response

Method	RMSE ↓	MSE gain ↑
Capability	1.172	0.0%
Direct pulse	**0.544**	**78.4%**
Operator	0.634	70.7%
Selector	0.554	77.7%

Direction

Method	Sign BA ↑
Capability	0.654
Direct pulse	0.535
Operator†	0.657
Full L-State	**0.664**

Action selection

Method	Regret ↓	Top-1 ↑
Capability	0.291	0.531
Direct pulse	0.215	0.623
Matched state	**0.212**	**0.648**
Full L-State†	0.217	0.636

†Prespecified readout. MSE gain uses capability as the reference. Selector denotes the action-wise pulse selector. Its RMSE pools all 540 scalar response components. Sign BA pools actions within each capability coordinate, computes balanced accuracy coordinate-wise, and macro-averages across coordinates. The matched-state reference uses aligned source-state responses and therefore a stronger information set than the deployment readouts.

B.4 SOURCE COVERAGE AND ACTION COORDINATES

Enumerate the source checkpoints as $\mathcal{S} = \{s_1, \dots, s_n\}$, where s_i belongs to family $m(s_i)$, and write $\delta_{s_i,u} = \beta_{m(s_i)} + \gamma_{m(s_i),u}$. Stack source operators and responses as

$$\mathcal{B}_S = \begin{bmatrix} B_{s_1} \\ \vdots \\ B_{s_n} \end{bmatrix}, \qquad y_S(u) = \mathcal{B}_S \bar{a}_u + v^{\mathrm{het}}_{S,u} + e_S,$$
$$v^{\mathrm{het}}_{S,u} = \begin{bmatrix} B_{s_1}\delta_{s_1,u} \\ \vdots \\ B_{s_n}\delta_{s_n,u} \end{bmatrix}, \qquad G_S = \mathcal{B}_S^\top \mathcal{B}_S = \sum_{s\in\mathcal{S}} B_s^\top B_s. \tag{19}$$

The source-coverage radius satisfies $\underline{\sigma}_S = \sigma_r(\mathcal{B}_S) = \sqrt{\lambda_{\min}(G_S)}$.

Proof of Theorem 3.3. With exact operators, $\widehat{a}_u - \bar{a}_u = \mathcal{B}_S^\dagger(v^{\mathrm{het}}_{S,u} + e_S)$, whose norm is at most $(H_{S,u} + \epsilon_S)/\underline{\sigma}_S$.

For estimated operators, Weyl's inequality gives $\sigma_r(\widehat{\mathcal{B}}_S) \geq \underline{\sigma}_S - \tau_S > 0$. Rewriting the source response as

$$y_S(u) = \widehat{\mathcal{B}}_S \bar{a}_u + (\mathcal{B}_S - \widehat{\mathcal{B}}_S)\bar{a}_u + v_{S,u}^{\mathrm{het}} + e_S$$

and left-multiplying by $\widehat{\mathcal{B}}_S^{\dagger}$ yields

$$\widehat{a}_u - \bar{a}_u = \widehat{\mathcal{B}}_S^{\dagger}\big[(\mathcal{B}_S - \widehat{\mathcal{B}}_S)\bar{a}_u + v_{S,u}^{\mathrm{het}} + e_S\big].$$

The pseudoinverse norm is at most $1/(\underline{\sigma}_S - \tau_S)$, proving Eq. (8). □

The source-diversity statement follows from $G_{S\cup\{\mathrm{new}\}} = G_S + B_{\mathrm{new}}^{\top} B_{\mathrm{new}}$ and Weyl monotonicity for positive semidefinite matrices. In particular,

$$\lambda_{\min}(G_{S\cup\{\mathrm{new}\}}) \geq \lambda_{\min}(G_S) + \sigma_r(B_{\mathrm{new}})^2.$$

If source block s has operator error at most β_s, the stacked perturbation satisfies $\tau_S \leq (\sum_s \beta_s^2)^{1/2}$. Applying Eq. (18) to each source block and taking a joint event therefore supplies the finite-repeat source term in Eq. (10).

For example, under $\underline{\sigma}_S \asymp \sqrt{n}$ and persistent $H_{S,u} \asymp \sqrt{n}$, the heterogeneity contribution to the displayed upper bound stays of constant order. The exact pooling bias is $G_S^{-1}\sum_{s\in\mathcal{S}} B_s^{\top} B_s \delta_{s,u}$; weighted centering of $\delta_{s,u}$ alone need not cancel this operator-weighted sum. More checkpoints can improve coverage and averaging of centered noise, while the pooling bias depends on family composition and operator geometry.

Under the same source conditions, including $\tau_S < \underline{\sigma}_S$, the ridge estimator in Eq. (24) also admits an explicit source bound. For $\lambda \geq 0$, let

$$\widehat{a}_{u,\lambda} = (\widehat{\mathcal{B}}_S^{\top}\widehat{\mathcal{B}}_S + \lambda I)^{-1}\widehat{\mathcal{B}}_S^{\top} y_S(u).$$

Let $\widehat{\sigma}_i$ denote the singular values of $\widehat{\mathcal{B}}_S$, and define

$$b_\lambda = \frac{\lambda}{\widehat{\sigma}_r^2 + \lambda}, \qquad \rho_\lambda = \max_i \frac{\widehat{\sigma}_i}{\widehat{\sigma}_i^2 + \lambda}.$$

For $\lambda > 0$, or for $\lambda = 0$ with full column rank,

$$\|\widehat{a}_{u,\lambda} - \bar{a}_u\|_2 \leq b_\lambda A + \rho_\lambda(\tau_S A + H_{S,u} + \epsilon_S). \tag{20}$$

Indeed, the ridge normal equations express the error as a regularization-bias term plus the ridge inverse applied to $(\mathcal{B}_S - \widehat{\mathcal{B}}_S)\bar{a}_u + v_{S,u}^{\mathrm{het}} + e_S$. Singular-value decomposition gives the two displayed coefficients; in particular, $\rho_\lambda \leq \min\{1/(\underline{\sigma}_S - \tau_S), 1/(2\sqrt{\lambda})\}$ for $\lambda > 0$.

B.5 Unseen-Family Transfer and Decisions

Proof of Theorem 3.4. By Theorem 3.2, $\|\widehat{B}_\star - B_\star\|_2 \leq \eta_\star/\sigma_r(Q_\star) =: \varepsilon_{B,\star}$. Let $\zeta_u = (\tau_S A + H_{S,u} + \epsilon_S)/(\underline{\sigma}_S - \tau_S)$ from Theorem 3.3. Writing $r_\star^h(u) = B_\star(\bar{a}_u + \delta_{\star,u}) + d_{\star,u}$ gives

$$\begin{aligned}\widehat{B}_\star\widehat{a}_u - r_\star^h(u) = (\widehat{B}_\star - B_\star)\bar{a}_u \\ + \widehat{B}_\star(\widehat{a}_u - \bar{a}_u) - B_\star\delta_{\star,u} - d_{\star,u}.\end{aligned}$$

Since $\|\widehat{B}_\star\|_2 \leq \|B_\star\|_2 + \varepsilon_{B,\star}$, taking norms and substituting $\|\bar{a}_u\|_2 \leq A$, $\|\delta_{\star,u}\|_2 \leq D_{\star,u}$, ζ_u, and $\varepsilon_{B,\star}$ gives Eq. (10). □

The radius in Theorem 3.4 is an analytical error decomposition. Using it as a prospective decision certificate requires independently valid bounds on the unseen-family residual and coordinate deviation. Source-family dispersion is a diagnostic proxy; certification for an unseen family additionally requires an assumption connecting that family to the observed dispersion.

For the ridge-regularized source estimator used by the operator readout, Theorem 3.4 holds after replacing the last fraction in Eq. (10) by the coordinate radius in Eq. (20). If scoring uses a realized response $\widetilde{r}_\star(u) = r_\star^h(u) + \omega_u$ with $\|\omega_u\|_2 \leq \kappa_u$, its total radius is $\overline{\Delta}_{\star,u} = \Delta_{\star,u} + \kappa_u$.

If $\epsilon_{\star,u}$ is instantiated by the smooth local remainder in Eq. (14), its curvature term uses $\|a_{\star,u}\|_2 \le A + D_{\star,u}$ and can therefore scale as $(A+D_{\star,u})^2$; $H_{S,u}$ accounts only for the linear pooling mismatch.

For completeness, let $\mathcal{C}_j \subseteq \{+,-\}$ contain the nonempty source-dead-zone-determinate classes of coordinate j, and let $M_{j,s}$ of $N_{j,s}$ cells in class s fail the sign margin. Over the valid coordinate set $\mathcal{J}$,

$$\mathrm{BA}_{\mathrm{macro}} \ge 1 - \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}}\frac{1}{|\mathcal{C}_j|}\sum_{s\in\mathcal{C}_j}\frac{M_{j,s}}{N_{j,s}}. \tag{21}$$

When $\rho_J = J(u^\star) - \min_u J(u) > 0$, Eq. (11) divided by ρ_J bounds normalized regret. For realized responses, define $\widetilde{J}(u) = hv^\top \widetilde{r}_\star(u) - \mathrm{cost}(u)$, $\widetilde{u}^\star \in \arg\max_u \widetilde{J}(u)$, and $\widetilde{\rho}_J = \max_u \widetilde{J}(u) - \min_u \widetilde{J}(u)$. Then

$$\widetilde{J}(\widetilde{u}^\star) - \widetilde{J}(\widehat{u}) \le h\|v\|_2\big(\overline{\Delta}_{\star,\widetilde{u}^\star} + \overline{\Delta}_{\star,\widehat{u}}\big), \tag{22}$$

and division by positive $\widetilde{\rho}_J$ gives realized normalized regret. The realized sign and exact-action conditions replace Δ by $\overline{\Delta}$.

Proof of Corollary 3.5. Equation (9) implies $|\widehat{r}_{\star,j}(u) - r^h_{\star,j}(u)| \le \Delta_{\star,u}$ for every coordinate. A response farther than this radius from zero keeps its sign, which proves the sign statement. Every class cell that clears the margin is therefore correct. For coordinate j and class s, recall is at least $1 - M_{j,s}/N_{j,s}$. Averaging first over the nonempty classes of each coordinate and then over valid coordinates proves Eq. (21).

For action selection, add and subtract predicted utilities and use optimality of $\widehat{u}$:

$$\begin{aligned} J(u^\star) - J(\widehat{u}) &\le |J(u^\star) - \widehat{J}(u^\star)| + |\widehat{J}(\widehat{u}) - J(\widehat{u})| \\ &\le h\|v\|_2(\Delta_{\star,u^\star} + \Delta_{\star,\widehat{u}}). \end{aligned}$$

Division by the positive utility range ρ_J gives the normalized-regret bound. For every $u \ne u^\star$, the predicted best-versus-u gap is at least the true gap minus $h\|v\|_2(\Delta_{\star,u^\star} + \Delta_{\star,u})$. The stated utility-margin condition makes all these differences positive, so the predicted maximizer is $u^\star$. Applying the same add-and-subtract argument to $\widetilde{J}$ and the simultaneous radii $\overline{\Delta}_{\star,u}$ proves Eq. (22); division by positive $\widetilde{\rho}_J$ gives its normalized form. For high-probability radii, take a joint event covering every source block, target pulse, candidate action, and scored coordinate before applying these deterministic arguments. A union bound over the finite index set constructs such an event from marginal tail bounds. □

C Complete Experimental Details

The direct pulse and full L-State readouts use the same ridge objective with $z(s) = p(s)$ and $z(s) = [c(s); p(s)]$, respectively. Let $\widetilde{z}(s)$ and $\widetilde{y}(s,u)$ denote source-standardized features and responses; an intercept b is unpenalized:

$$(\widehat{W}_{u,\alpha}, \widehat{b}_u) = \arg\min_{W,b} \sum_{s\in\mathcal{S}_{\mathrm{src}}} \|W^\top \widetilde{z}(s) + b - \widetilde{y}(s,u)\|_2^2 + \alpha\|W\|_F^2, \tag{23}$$

Each feature and response coordinate uses the mean and population standard deviation of the pooled source training rows, with a unit divisor for constant coordinates. These transformations are refitted inside each source-trajectory validation fold, then on all source rows, and applied unchanged to the held-out family; predictions are transformed back to response units. Numerically rank-deficient feature designs use a source-fitted 99%-energy singular-vector projection. Thus the static loss features are source-standardized across tokenizers, while their definition remains per token; this scaling does not establish tokenizer-invariant capability units.

The structured operator objective used in Section 3.6 is

$$\widehat{a}_u = \arg\min_a \sum_{s\in\mathcal{S}_{\mathrm{src}}} \|R(s)a - y(s,u)\|_2^2 + \lambda\|a\|_2^2, \qquad \widehat{y}_{\mathrm{op}}(s,u) = R(s)\widehat{a}_u. \tag{24}$$

C.1 FAMILY-LEVEL GATING AND BASELINES

In development, each outer fold trains on two model families and tests on the third. Source-trajectory inner cross-validation selects hyperparameters, and predictions are fixed before the held-out family's target responses enter evaluation. The complete comparison includes a source-mean reference, CAPABILITY, a direct pulse readout, scalar training statistics, a trajectory/index-matched source state, and the OPERATOR readout. TARGET EARLY provides a separate target-specific reference.

Each sealed stage freezes its reporting rule, 540 core response predictions, and 810 action choices before target training. Granite additionally freezes 216 selector-extension rows copied from the chosen direct/operator predictions. Five core methods use 27 states, four actions, five coordinates, and six utilities. Capability is the common reference; matched state averages development responses at the same seed, state index, and action.

C.2 STATE GENERATION

Each trajectory starts from the same base model with a newly initialized zero-output rank-4 LoRA (Hu et al., 2022) adapter. Four synthetic generators—sequence reversal, lexicon classification, date normalization, and template paraphrase—appear twice in a balanced eight-episode order. States $t00$ through $t08$ are saved. Three fixed trajectory seeds give nine states per trajectory and 27 per family. The three development families contribute 81 states; GLM and Granite each add three fresh trajectories and 27 states, for 135 states in total. GLM's science inversion motivates the development-only selector, which is frozen before Granite target training.

C.3 SYNTHETIC TASK CONSTRUCTION

The supplementary generator specifies the complete distributions and literal prompt strings. A single Python pseudorandom stream with seed 20260829 generates the tasks in the order below: 1,024 rows for each state generator, then 256 rows for each probe. Row index i starts at zero within each task. Define the ordered symbol list `[vek,lun,zor,pim,taf,gex,rud,bov]` and colors `[amber,cobalt,ivory,jade]`. Random integers below are uniform with inclusive endpoints; sampling permits replacement unless specified otherwise.

State generators. Sequence reversal samples five integers in $[100, 998]$, prefixes each with `rv`, and asks to reverse the space-separated list. Lexicon classification shuffles the eight symbols, assigns colors cyclically in that shuffled order, and queries the symbol at position $i \bmod 8$. Date normalization uses year $2100+(i \bmod 70)$, month $1+(i \bmod 12)$, and day $1+(7i \bmod 28)$; the input alternates by $i \bmod 3$ between `DD.MM.YYYY`, `YYYY/MM/DD`, and `Mmm-Ddd-Yyyyy`. The answer is `YYYY-MM-DD`. Template paraphrase substitutes `unit` $(i \bmod 97)$ and `crystal` $(11i \bmod 89)$ into the fixed sentence shown below. The remaining wording is fixed.

Standard probes. Schema mapping samples four distinct letters from `abcdef`, prefixes each key with `k`, and assigns values `v`n with $n \in [1000, 9998]$. The prompt preserves sampled field order; the answer is JSON sorted by key with commas and no spaces. Rule chaining samples three entity identifiers `e`n, $n \in [100, 998]$, and applies the fixed two-link rule in the example. Symbol rewriting samples six symbols from the ordered list, replaces them by `A` through `H` in list order, and concatenates the outputs; the prompt renders the mapping as a Python dictionary in that order. Table aggregation samples five integers in $[3, 96]$ for `z0` through `z4`, then queries positions $i \bmod 4$ and $(i + 2) \bmod 5$. These definitions, the fixed seed, and generation order recover the example pools exactly. The supplement includes episode selection and ordering code as well as the frozen data configuration.

Prompt and completion examples. The following are the first generated pool rows. Bracketed legends, mappings, records, queried entities, and numerical slots vary according to the rules above; the instruction text stays fixed. Chat tasks place the prompt in the user role and the completion in the assistant role through the checkpoint's frozen chat template. The code masks prompt tokens and limits sequences to 512 tokens. The tokenizer determines the token count for the same text. The Granite evaluation prepends the frozen system message `Knowledge Cutoff Date: April 2024.`

`Today's Date: August 31, 2026.`
`You are Granite, developed by IBM. You are a helpful AI assistant.`
The three lines are separated by newline characters; the same encoding is used before and after each branch.

Sequence reversal. Prompt: `Reverse this synthetic rune sequence exactly: rv717 rv522 rv449 rv772 rv768`
Completion: `rv768 rv772 rv449 rv522 rv717.`

Lexicon classification. Prompt: `Using the fictional lexicon [bov=amber; lun=cobalt; taf=ivory; vek=jade; zor=amber; pim=cobalt; gex=ivory; rud=jade], return only the class of bov.`
Completion: `amber.`

Date normalization. Prompt: `Normalize invented-calendar case N0000, date 01.01.2100, to YYYY-MM-DD.`
Completion: `2100-01-01.`

Template paraphrase. Prompt: `Paraphrase while preserving slots SUBJECT=unit0; OBJECT=crystal0: 'SUBJECT catalogued OBJECT at noon.'`
Completion: `At noon, unit0 entered crystal0 in the catalogue.`

Schema mapping. Prompt: `Encode and key-sort this Kestrel record as minified JSON: kd: v4303 | ka: v6427 | kf: v4803 | kb: v4596`
Completion: `{"ka":"v6427","kb":"v4596","kd":"v4303","kf":"v4803"}`.

Rule chaining. Prompt: `In the Nacre rules: e722 feeds e576; e576 marks e800. Rule: anything feeding a marker selects what it marks. Return the selected entity for e722.`
Completion: `e800.`

Symbol rewriting. Prompt: `Apply the Quartz rewrite {'vek': 'A', 'lun': 'B', 'zor': 'C', 'pim': 'D', 'taf': 'E', 'gex': 'F', 'rud': 'G', 'bov': 'H'} to vek+pim+rud+rud+gex+taf, then concatenate the output symbols.`
Completion: `ADGGFE.`

Table aggregation. Prompt: `From the Umber table [z0=79, z1=4, z2=87, z3=5, z4=19], return z0+z2 as one integer.`
Completion: `166.`

C.4 Primary Training Episode

The base model is frozen. LoRA rank is 4 with scaling 8, dropout 0, and query and value projection targets. The GLM checkpoint is specifically GLM-4-9B-0414, loaded as `Glm4ForCausalLM`. Its separate `q_proj` and `v_proj` modules receive the same adapters as the other families; the loader checks that both module names exist before injection. Training uses BF16, eager attention, maximum length 512, micro-batch 1, gradient accumulation 4, AdamW learning rate 10^{-4}, betas $(0.9, 0.999)$, epsilon 10^{-8}, no weight decay, and global gradient clipping at 1.0. Eight optimizer steps consume 32 examples. Every branch creates new optimizer moments and restores the source adapter afterward.

C.5 Response Evaluation

Each capability is the mean negative completion-token NLL after per-token clipping to $[0, 20]$, so larger is better. Before and after evaluations use the same 32 held-out examples and tokenizer outputs. The primary response is the paired per-example difference. The main state and readouts use this paired quantity; generated-answer diagnostics are stored as a separate trace. GSM8K evaluation scores the extracted final answer, while its training completion includes the derivation. MBPP scores code completion tokens, SciQ the option letter and answer, BoolQ Yes/No, and WikiText-2 causal

text tokens. These likelihood endpoints and generated-answer accuracy measure distinct outcomes, so response MSE gains quantify the former.

C.6 Reliability Thresholds and Response Slices

For capability coordinate j and pulse q, let $r^{(b)}_{s,q,j}$ be the response for calibration state s and order repeat $b \in \{0, 1, 2\}$, with repeat mean $\bar{r}_{s,q,j}$. Over $N = 18$ aligned calibration states from Qwen, Mistral, and OLMo, define

$$s^2_{q,j} = \frac{1}{N(3-1)} \sum_{s=1}^{N} \sum_{b=0}^{2} (r^{(b)}_{s,q,j} - \bar{r}_{s,q,j})^2, \qquad t_j = \text{median}_{q=1,\ldots,4}(1.96 s_{q,j}). \tag{25}$$

In mathematics/code/science/reading/general order, the frozen thresholds are $(0.09130661, 0.01883772, 0.01401346, 0.09772611, 0.04695581)$. Observed components with $|y_j| > t_j$ enter sign scoring; predicted signs use zero. The factor 1.96 scales within-state repeat variation rather than the standard error of a repeat mean. The development suite pools pulse-only reliability across its three families, including the family held out by each prediction fold. The GLM and Granite stages reuse these same thresholds, fixed before either held-out family's feature collection; their target-action labels and pulse repeats do not calibrate the thresholds. A post-hoc decomposition of the already frozen predictions reports source-standardized RMSE separately inside and outside the sign dead zone (Table 5). The same fixed mask is used for every method. These descriptive slices retain the original source scaling and introduce no refitting or threshold selection. Both direct and operator readouts improve RMSE over capability on the determinate components in both sealed families, showing predictive value for changes larger than the reliability threshold.

C.7 Leakage Controls

The data registry records source IDs, processed files, and roles. Exact cross-role text intersections are zero; source indices for public train and evaluation splits are disjoint; synthetic pulse-to-target 13-gram Jaccard is below 0.01. The analysis reads validated completed results, and target responses are evaluated after the outer-fold predictions are fixed.

D Development Results

D.1 Aligned-Pulse Permutation

For every outer fold and permutation, we shuffle whole pulse-response matrices across states within each source family, rebuild the complete 25-dimensional input, repeat source-trajectory hyperparameter selection, refit, and score the unchanged target family. Table 10 reports the one-sided family results. We report raw per-fold one-sided p-values; the preregistered comparison uses each fold's null 95th percentile.

D.2 Source-Fit Blend Weights

Weights below multiply the learned readout; one minus the weight multiplies the operator readout. Reciprocal source-family fits select the weights, while the pooled three-development-family pulse reliability fixes the sign dead zones (Appendix C.6); held-out target-action labels remain sealed. This development-suite interpolation raises sign balanced accuracy from 0.589 to 0.643 and reduces RMSE from 8.293 to 7.876. The first GLM seal uses the prespecified operator/full-state readout choices summarized in the main paper.

E Prospective GLM Details

E.1 Prediction Timing and Complete Comparison

The prospective evaluation uses the rule learned from the 81 development states. After measuring four target-independent pulse responses at each of 27 GLM states, we fix 540 response predictions

Table 5: Frozen-prediction RMSE by response magnitude relative to the reliability threshold. D: determinate; I: inside the dead zone. Counts are scalar components per method.

	GLM		Granite	
Readout	D ($n = 184$)	I ($n = 356$)	D ($n = 257$)	I ($n = 283$)
Capability	1.5826	1.1939	1.0810	1.2481
Direct pulse	1.0542	0.4403	0.6651	0.4033
Matched state	1.1695	0.7944	0.7481	0.6359
Operator	1.0592	0.1033	0.8900	0.2205
Full L-State	1.1734	0.6004	0.7591	0.6002

Table 6: Complete three-family leave-one-family-out development comparison. The direct pulse and operator methods read the same L-State pulse block. RMSE is source-standardized; regret is realized normalized regret under six fixed utilities. The directional blend is evaluated in the development suite. Target early uses target-specific responses after two update steps.

Method	RMSE ↓	Cosine ↑	Sign BA ↑	Regret ↓	Top-1 ↑
Source mean	7.895	0.624	0.479	0.268	**0.574**
Capability	9.987	0.504	0.602	0.472	0.309
Direct pulse	**7.772**	0.620	0.531	0.299	0.500
Scalar statistics	8.738	0.591	0.547	0.365	0.449
Matched state	7.913	0.607	0.491	**0.232**	0.566
Operator	7.777	0.672	**0.675**	0.303	0.350
Full L-State	8.293	**0.734**	0.589	0.248	0.541
Target early (2 steps)	8.067	0.488	0.671	0.283	0.416
Directional blend	7.876	0.628	0.643	0.292	0.422

Table 7: Full L-State readout results by held-out family.

Family	Full L-State	Capability	MSE gain	Sign BA
Qwen	0.926	2.947	0.901	0.787
Mistral	13.866	15.117	0.159	0.610
OLMo	3.631	7.876	0.787	0.500

Table 8: Family heterogeneity of sign and realized regret.

Family	Capability BA	Operator BA	Capability R	L-State R	Matched R
Qwen	0.829	0.743	0.497	0.375	0.352
Mistral	0.617	0.752	0.276	0.143	0.235
OLMo	0.561	0.521	0.643	0.225	0.110

Table 9: Pulse-count ablation for the learned readout.

k	RMSE	MSE gain	Cosine	Sign BA	Regret
1	9.108	0.168	0.670	0.604	0.300
2	9.614	0.073	0.650	0.615	0.259
3	8.358	0.300	0.724	0.598	0.262
4	8.293	0.311	0.734	0.589	0.248

Table 10: Aligned-pulse permutation controls with source refitting.

Family	Aligned gain	Null 95th	p
Qwen	0.901	0.873	0.0010
Mistral	0.159	0.135	0.0330
OLMo	0.787	0.870	0.6014

Table 11: Source-fit directional blend weights by held-out development family.

Held out	Math	Code	Science	Reading	General
Qwen	0.00	0.80	0.05	0.00	0.50
Mistral	0.05	0.00	0.30	0.75	0.05
OLMo	0.30	0.85	0.00	0.00	0.10

and 810 utility-conditioned action decisions before any target training. Scoring begins after all 108 GLM target branches finish. Table 12 gives all five frozen methods.

E.2 Response Slices

The determinate direction set contains 169 positive and 15 negative responses. The general coordinate falls inside its source-frozen dead zone for all 108 records, so its direction entry is shown as a dash.

E.3 Action and Trajectory Slices

Whole-trajectory resampling gives percentile ranges of $[78.1\%, 78.5\%]$ for operator MSE reduction, $[0.705, 0.860]$ for its sign BA, $[32.2, 52.3]$ points for its sign gain, and $[25.0\%, 29.9\%]$ for full-state regret reduction. These intervals summarize variation across the three prospective GLM trajectories.

F Mechanism Diagnosis and Action-Wise Response Selection

F.1 Five-Repeat Science Intervention

On the 28 determinate GLM science-action coordinates, operator sign BA is zero and direct BA is 1.0. Four additional target realizations preserve the target sign in all 27 states; source-fit coordinate cosines span -0.555 to 0.322. The target responses retain their signs across repeats, whereas the source-fitted coordinates vary across development-family folds. Using only development-family folds, the selector chooses operator for mathematics/reading and direct for code/science, then freezes this mapping before Granite labels open.

Four additional independent science-action branches are run from each of the 27 GLM states, adding 108 branches. Each branch restores its initial state adapter. The original target responses and the capability, direct, and operator predictions remain unchanged. Table 16 evaluates these predictions against the original response and the five-response mean and median.

Across the 27 states, the science-coordinate response has mean within-state standard deviation 0.00446 and unanimous five-repeat sign in 27/27 states. The source science-action coordinate uses $\lambda = 1000$ and has norm 0.0539; its held-out-family fits have norms 0.0588–1.108 and pairwise cosine -0.555–0.322.

F.2 Development-Fitted Readout Selection

The direct readout is selected after a strict held-out-family win in at least two of the three development folds; the remaining actions retain the operator readout. Table 17 records the frozen mapping. The direction selector is retained as a diagnostic; the Granite direction analysis uses the prespecified operator readout.

Table 12: Complete prospective GLM comparison and response slices. Direction uses the 184 scalar responses selected by source-frozen dead zones; continuous RMSE retains all 540 components. The direct pulse and operator methods read the same L-State pulse block.

Method	RMSE ↓	MSE gain ↑	Cosine ↑	Sign BA ↑	Regret ↓	Top-1 ↑
Capability	1.339	0.000	0.431	0.366	0.592	0.340
Direct pulse	0.712	0.718	**0.552**	0.571	0.451	0.389
Matched state	0.939	0.508	0.535	0.569	**0.410**	**0.432**
Operator	**0.624**	**0.783**	0.421	**0.754**	0.513	0.352
Full L-State	0.841	0.606	0.471	0.359	0.434	0.395

Response and direction by target action

Action	Cap. RMSE	Direct RMSE	Oper. RMSE	Cap. BA	Direct BA	Oper. BA	Det.
u_{math}	1.533	0.541	0.085	0.833	0.833	0.833	45
u_{code}	1.214	0.571	0.454	0.479	0.500	0.850	48
u_{science}	0.966	0.613	0.825	1.000	1.000	0.000	28
u_{reading}	1.554	1.015	0.815	0.167	0.438	0.641	63

Response and direction by capability coordinate

Coordinate	Cap. RMSE	Direct RMSE	Oper. RMSE	Cap. BA	Direct BA	Oper. BA	Det.
Math	1.366	0.720	0.216	0.071	0.071	0.721	19
Code	1.553	0.546	0.499	1.000	1.000	1.000	52
Science	1.244	0.984	1.024	0.392	0.399	0.516	86
Reading	1.686	0.852	0.772	0.000	0.815	0.778	27
General	0.545	0.146	0.073	–	–	–	0

Table 13: Prospective action selection by utility. Cells report regret/top-1.

Utility	Capability	Full L-State	Matched state
Math	0.987/0.000	**0.398/0.074**	0.539/0.037
Code	0.000/1.000	0.000/1.000	0.000/1.000
Science	0.000/1.000	0.000/1.000	0.000/1.000
Reading	0.988/0.000	0.985/0.000	**0.797/0.185**
General	0.749/0.037	0.474/0.296	**0.375/0.370**
Balanced	0.827/0.000	**0.745/0.000**	0.749/0.000

Table 14: Prospective metric estimates by GLM trajectory.

Seed	Operator MSE gain	Operator BA	BA gain	L-State regret	Regret reduction
20260901	0.781	0.706	0.322	0.434	0.254
20260902	0.785	0.735	0.372	0.409	0.299
20260903	0.783	0.860	0.523	0.457	0.250

On Granite, the ACTION-WISE PULSE SELECTOR reaches pooled and action-macro RMSE of 0.5537 and 0.5479. Its pooled and action-macro sign BA are 0.6496 and 0.6470. The source-only direction diagnostic has pooled BA 0.5689 and action-macro BA 0.6558.

Table 15: Whole-trajectory bootstrap summaries with 10,000 valid resamples.

Quantity	Estimate	2.5%	97.5%
Operator MSE gain	0.783	0.781	0.785
Operator sign BA	0.754	0.705	0.860
Operator sign BA gain	0.388	0.322	0.523
L-State regret reduction	0.267	0.250	0.299

Table 16: GLM science-action metrics under repeat-based truth summaries. Each cell is source-standardized RMSE/sign BA.

Truth summary	Capability	Direct pulse	Operator
Original	0.9664/1.000	0.6132/1.000	0.8246/0.000
Five-repeat mean	0.9667/1.000	0.6115/1.000	0.8245/0.000
Five-repeat median	0.9643/1.000	0.6111/1.000	0.8243/0.000

Table 17: Action-wise pulse selector. "Wins" counts strict direct-readout wins over the operator readout among three held-out development families.

Action	Response	Wins	Direction	Wins
Math	Operator	0/3	Operator	0/3
Code	Direct	3/3	Operator	1/3
Science	Direct	2/3	Direct	3/3
Reading	Operator	1/3	Direct	2/3

G Five-Family Semantic-Coordinate Analysis

Table 18 tests the structure imposed by the operator after all five families' labels open. Pooling error is smallest for mathematics and largest for reading; science has the lowest shared-signal fraction, and reading's mean coordinate cosine is negative. At level 0.05, exact sharing is rejected for code, science, and reading; mathematics remains unresolved. All four $D_u^{\mathrm{DZ}} > 1$, so none meets the retrospective equivalence criterion. Joint tests detect family effects, total deviation, and the family–action interaction increment (all $p \leq 3.05 \times 10^{-4}$). Shared, family-main, and full family–action models have held-trajectory RMSE 0.824, 0.802, and 0.750. The full model reduces MSE by 17.3% overall and 48.0% for science. Appendix G gives decomposition and sensitivity checks.

G.1 Analysis Overview

Only after all five families' labels open, full-rank frozen action matrices recover pulse operators with prediction replay error below 10^{-9}. One development-frozen response scale, complete trajectory blocks, and exact sign flips over 15 family–trajectory clusters compare the five observed families under their fixed measurements and target realization. Sealed predictions remain unchanged. The following subsections give recovery, conditioning, and sensitivity details.

G.2 Operator Recovery and Coordinate Estimation

The analysis uses 540 previously scored operator family–action–state rows: five families, four actions, and 27 states per family. The frozen four-action coordinate matrix is full rank in each of the five family folds. Their condition numbers are 23.4 for Qwen, 1244.8 for Mistral, 34.7 for OLMo, and 33.6 for both GLM and Granite. Algebraic operator recovery replays every frozen operator-readout prediction with maximum absolute error below 10^{-9}. Re-fitting the three development-family science coordinates from the recovered operators agrees with the original direct fit to 3.3×10^{-15}, providing a check beyond prediction replay.

Table 18: **Testing the pooled action coordinate.** D_u: relative pooling error; S_u: shared-signal fraction; D_u^{DZ}: error in frozen-dead-zone units. Brackets: 95% trajectory-block percentile intervals conditional on the five observed families and selected ridge.

Action	D_u [95% interval]	S_u	D_u^{DZ} [95% interval]
Math	.342 [.311,.370]	.648	2.468 [2.252,2.710]
Code	.356 [.326,.394]	.800	2.620 [2.454,2.815]
Science	.633 [.605,.672]	.269	4.417 [4.324,4.523]
Reading	1.038 [.976,1.105]	.404	2.837 [2.714,2.998]

The ridge grid spans zero through 10 and is refined between 0.1 and 2. Holding out a complete trajectory in every family–action cell selects relative ridge 0.7. Each absolute cell penalty equals 0.7 times the mean eigenvalue of that cell's design Gram. All reported response-scale quantities use the common three-development-family scale. The family-specific outer-fold scale and every candidate ridge remain in the sensitivity results.

For this retrospective decomposition, the reference set is the five observed families with $\pi_m = 1/5$; write its mean as $\bar{a}_u^{(5)}$ to distinguish it from the prospective source-reference coordinate, and suppress the superscript below. Let W_u be the inverse-squared response scale frozen on the three development families, and let Z analogously use the frozen reliability dead zones. For N family–state cells and d response coordinates, define

$$
\begin{aligned}
D_u^2 &= \frac{1}{Nd} \sum_{m,s} \left\| W_u^{1/2} R_m(s)(\beta_m + \gamma_{m,u}) \right\|_2^2, \\
(D_u^{\mathrm{DZ}})^2 &= \frac{1}{Nd} \sum_{m,s} \left\| Z^{1/2} R_m(s)(\beta_m + \gamma_{m,u}) \right\|_2^2, \\
S_u &= \frac{L_u^2}{L_u^2 + D_u^2}, \qquad L_u^2 = \frac{1}{Nd} \sum_{m,s} \| W_u^{1/2} R_m(s) \bar{a}_u \|_2^2.
\end{aligned}
\tag{26}
$$

Thus D_u is common-scale pooling error and S_u is the shared-signal fraction. Exact tests enumerate all 2^{15} restricted-residual cluster sign patterns and use Holm correction across actions. The separate diagnostic equivalence test is $H_0 : D_u^{\mathrm{DZ}} \geq 1$ against $H_1 : D_u^{\mathrm{DZ}} < 1$; the dead zones were frozen prospectively, but their use as a sharedness margin was chosen retrospectively. Tests and predictive comparisons use unpenalized least squares; ridge 0.7 applies only to coordinate-stability summaries.

G.3 Difference, Interaction, and Predictive Diagnostics

The action-specific restricted-residual sign-flip values before/after Holm adjustment are .0874/.0874 for mathematics, .00165/.00494 for code, 6.10×10^{-5}/.000244 for science, and .0239/.0477 for reading. The joint family-main, total-deviation, and interaction-increment values are 6.10×10^{-5}, 1.83×10^{-4}, and 3.05×10^{-4}. These are restricted-residual tests over the 15 observed family–trajectory clusters, assuming approximate residual sign symmetry.

Compared with the shared model, the full model's descriptive held-trajectory MSE improvement is 17.3% overall, 0.3% for mathematics, 18.8% for code, 48.0% for science, and 1.8% for reading. Descriptive resampled ranges summarize variability across the three overlapping training folds.

Deleting one family at a time gives mean-cosine ranges .628–.781 for mathematics, .573–.700 for code, $-.121$–.486 for science, and $-.136$–.167 for reading. The corresponding D_u ranges are .163–.371, .285–.427, .500–.752, and .555–1.333. The science and reading effects persist after removing Mistral, while their magnitude depends on the particular observed family set.

Table 19: Direct five-family coordinate stability. Intervals are 10,000 trajectory-block percentile intervals conditional on the five observed families and selected ridge. "DZ exceed" is the fraction of response cells in which pooling error exceeds the frozen reliability dead zone.

Action	Cosine min/mean	Norm ratio	D_u (95%)	S_u	D_u^{DZ} (95%)	DZ exceed
Math	.315/.696	3.55	.342 [.311,.370]	.648	2.468 [2.252,2.710]	24.3%
Code	.448/.653	2.43	.356 [.326,.394]	.800	2.620 [2.454,2.815]	39.7%
Science	−.770/.057	11.20	.633 [.605,.672]	.269	4.417 [4.324,4.523]	40.1%
Reading	−.732/−.017	19.43	1.038 [.976,1.105]	.404	2.837 [2.714,2.998]	32.3%

Table 20: Retrospective three-fold held-trajectory RMSE under the common response scale. The family-aware models are fitted to opened labels from each of the five observed families.

Coordinate model	All	Math	Code	Science	Reading
Shared action	.824	.757	.887	.800	.848
+ family main effect	.802	.794	.863	.720	.823
+ family–action interaction	.750	.756	.799	.577	.840

H Prospective Granite Details

H.1 Complete Core Comparison

The Granite evaluation fixes the prompt and tokenization, the 540-row core predictions, the 810-row action decisions, and the 216-row selector extension before target labels open. All 108 target branches complete; independent analysis reruns reproduce the core and selector-extension outputs.

Whole-trajectory resampling gives $[69.9\%, 71.1\%]$ for the operator MSE reduction, $[0.608, 0.658]$ for its sign BA, $[-6.15, 8.23]$ points for its sign gain, and $[15.7\%, 40.1\%]$ for full-state regret reduction. The sealed Granite result reproduces large response and action gains while replacing the GLM science-action inversion with non-degenerate direction estimates.

I Pulse Identification and Cross-Duration Robustness

I.1 Pulse Identification and Local-Dynamics Checks

Theorem B.1 requires a stable local response measurement, and Theorems 3.2–B.2 connect operator recovery to the pulse basis. Pulse measurement is stable: the median intraclass correlation is 0.967 and the median signal-to-noise ratio is 29.2, with 19 of 20 reliability cells at ICC 0.75 or above. The development and GLM results show that direction accuracy also depends on how these L-State measurements are read.

Pulse-count ablation shows that three to four directions carry the strongest continuous signal. One, two, three, and four pulses yield RMSE 9.108, 9.614, 8.358, and 8.293, respectively; action regret is 0.300, 0.259, 0.262, and 0.248. The best pulse count depends on the metric: sign balanced accuracy peaks with two pulses at 0.615, while RMSE and regret attain their best values with four. Three or four pulses give the strongest response predictions; two work best for direction.

Finally, controlled pulse mixtures have median relative residual 0.342; changing exposure from eight steps to four or sixteen gives median relative residual 0.747; and a fixed-step amplitude scan gives 0.757 away from the reference amplitude. The smaller mixture residual and the duration response quantify the finite-step geometry induced by the smooth-factorization result.

The cross-duration study reuses nine GLM states and evaluates four-, eight-, and sixteen-step pulse and target exposures. The operator readout has the best aggregate raw response RMSE and sign BA

Table 21: Complete sealed Granite comparison. RMSE is source-standardized; direction is coordinate-macro balanced accuracy. The direct pulse and operator methods read the same L-State pulse block.

Method	RMSE	MSE gain	Cosine	Sign BA	Regret	Regret red.	Top-1
Capability	1.1716	0.0000	**0.7382**	0.6537	0.2906	0.0000	0.5309
Direct pulse	**0.5439**	**0.7845**	0.3155	0.5354	0.2149	0.2604	0.6235
Matched state	0.6916	0.6516	0.4710	0.5351	**0.2119**	**0.2707**	**0.6481**
Operator	0.6344	0.7068	0.1725	0.6566	0.2530	0.1296	0.5926
Full L-State	0.6805	0.6627	0.5079	**0.6640**	0.2166	0.2545	0.6358

Table 22: Granite operator-readout response and direction by target action.

Action	RMSE	Sign BA	Determinate
Math	0.5963	0.6958	79
Code	0.8878	0.8252	89
Science	0.2860	0.4256	43
Reading	0.6198	0.6420	46

Table 23: Granite whole-trajectory bootstrap over the three observed trajectories (10,000 valid resamples; seed 20260919). Estimate is the aggregate point estimate; interval columns are percentile bounds.

Quantity	Estimate	2.5%	97.5%
Operator MSE gain	0.7068	0.6990	0.7110
Operator sign BA	0.6566	0.6081	0.6584
Operator sign BA gain	0.0028	−0.0615	0.0823
L-State regret reduction	0.2545	0.1565	0.4012

among the five core methods at every duration. The full L-STATE readout reduces source-standardized RMSE against capability by 45.3%, 37.3%, and 25.1%, and reduces action regret by 21.8%, 22.7%, and 29.3% at four, eight, and sixteen steps. Both improvements hold in all nine duration-by-trajectory slices. Retrospectively fitted operator action coordinates retain cosine 0.952–0.997 at four steps and 0.973–0.992 at sixteen steps relative to the eight-step coordinate. These within-GLM comparisons quantify coordinate stability across training durations. The prespecified metrics and coordinate-stability diagnostics appear below; the supplementary results include the complete five-method raw-primary, per-step, and trajectory tables.

Extending both pulse and target horizons tests response measurement during longer optimization. At matched 32/64/128-step horizons, operator MSE reductions are 53.9%, 54.1%, and 54.2%, with the lowest RMSE and highest sign BA among the five readouts. All source fits remain fixed. For the fixed-eight-step measurement, the four probes cost 32 optimizer steps, versus 128 steps to run four 32-step target trials, a 75% reduction in training steps for obtaining the measurements. This accounting excludes evaluation and fitting time. Appendices I–J give complete metrics, training settings, and horizon-dependent training-step costs.

I.2 REGISTERED DURATION EXPERIMENT

The duration experiment evaluates nine GLM states under four-, eight-, and sixteen-step pulse and target exposures. Raw capability difference is the primary response; per-step response is a separately reported diagnostic. The eight-step branches are reused read-only, 144 new branches cover the other two durations, and every adapter is restored.

At four, eight, and sixteen steps, capability/full-L-State source-standardized RMSE is 1.2801/0.7000, 1.3147/0.8240, and 1.4574/1.0922, respectively, yielding the reported 45.3%, 37.3%, and 25.1% reductions.

Table 24: **Training-response prediction across horizons.** Nine GLM states, three trajectories; steps count from the same initial state. Original source fits and training examples are retained. Scores use the original evaluation set, later used for selection; independent final-set scores appear in Table 39. Fixed-eight-step rows are post-hoc, without refitting or horizon rescaling. Bold: lowest core RMSE per row. The upper block extends both pulse and target horizons; the lower fixes eight-step pulses. BA compares capability and operator direction. The last column gives the operator's relative MSE reduction against capability alone.

Steps		Raw response RMSE ↓					Sign BA ↑		MSE reduction
Pulse	Target	Capability	Direct	Matched	Operator	Full state	Capability	Operator	Operator ↑
32	32	0.4742	0.3566	0.3817	**0.3220**	0.3581	0.4754	0.6394	53.9%
64	64	0.5395	0.4375	0.4680	**0.3656**	0.4294	0.4532	0.6544	54.1%
128	128	0.8427	0.7198	0.7782	**0.5704**	0.7287	0.4099	0.6246	54.2%
8	16	0.4861	0.3111	0.3786	**0.2943**	0.3728	0.3814	0.7490	63.3%
8	32	0.4742	**0.3249**	0.3817	0.3421	0.3775	0.4754	0.7617	47.9%

Table 25: Duration results for the prespecified readouts. Response cells report raw RMSE/sign BA for capability and the operator readout; action cells report normalized regret/top-1 for capability and the full L-State readout.

Steps	Cap. response	Oper. response	Cap. action	L-State action
4	.4428/.3564	**.1012/.6793**	.5592/.3333	.4375/.4259
8	.4529/.3828	**.1875/.6623**	.5771/.3333	.4460/.3889
16	.4861/.3814	**.2878/.7878**	.6063/.3333	.4285/.3704

Across all nine duration-by-trajectory slices, the full L-STATE readout has lower response RMSE and lower action regret than capability. Duration-specific operator diagnostic coordinates have cosine 0.952–0.997 at four steps and 0.973–0.992 at sixteen steps relative to the eight-step coordinate.

I.3 MATCHED PULSE AND TARGET HORIZONS UP TO 128 STEPS

The extension uses the same nine GLM checkpoints: states $t00$, $t04$, and $t08$ from each of three trajectories. Pulse and target branches both run for $h \in \{32, 64, 128\}$ steps, with each branch restored to its original checkpoint and optimizer moments reset. The original eight-step source-fitted readouts, feature transformations, scoring rules, and sign dead zones remain fixed. Each branch repeats its original ordered 32 training examples four, eight, or sixteen times, respectively. There is no target-family refitting. The extension completes 108 pulse branches before freezing all predictions and starting the 108 target branches, giving 216 new branches and 16,128 optimizer steps in total. The original eight-step results in Table 25 are reused as the anchor.

At each horizon, each readout supplies 36 state–action response vectors with five coordinates and 54 utility-conditioned action decisions. The nine states and six decisions per state are correlated; the three trajectories are the replication units. Tables 26 and 27 report descriptive point estimates, with the trajectory breakdown retained in the accompanying results. The prediction method combines the measured L-State pulse operator with source-calibrated action coordinates, as defined in Section 3.6. CAPABILITY is the common reference. All five original readouts and their utility-conditioned decision metrics are included for comparison.

The operator reduces raw-response MSE relative to capability by 53.9%, 54.1%, and 54.2% at 32, 64, and 128 steps, respectively, and has the lowest raw RMSE and highest sign BA among the five readouts at each horizon. At 128 steps, it reaches raw RMSE 0.5704 and sign BA 0.6246, compared with capability's 0.8427 and 0.4099. The source-calibrated L-State pulses predict both response magnitude and direction through 128 steps of training on the same repeated examples.

Table 26: Matched pulse and target horizons on nine GLM states. Response error uses raw capability-change units; sign BA retains the source-frozen dead zones. Regret and top-1 use the original six utilities. Bold marks the best point estimate among all five readouts within each horizon and metric. The response-prediction method is the source-calibrated L-State operator; the full-state ridge map is a separate readout comparison.

Steps	Readout	Raw RMSE ↓	Sign BA ↑	Regret ↓	Top-1 ↑
32	CAPABILITY	0.4742	0.4754	0.5603	0.3889
	Direct pulse	0.3566	0.3584	0.4117	0.4074
	Matched source state	0.3817	0.5769	0.3728	**0.4444**
	OPERATOR	**0.3220**	**0.6394**	0.3842	0.2778
	FULL L-STATE	0.3581	0.3584	**0.3690**	0.4074
64	CAPABILITY	0.5395	0.4532	0.5189	0.3889
	Direct pulse	0.4375	0.3930	0.4527	0.3704
	Matched source state	0.4680	0.6350	0.4397	0.4630
	OPERATOR	**0.3656**	**0.6544**	0.4498	0.4259
	FULL L-STATE	0.4294	0.3730	**0.3963**	**0.4815**
128	CAPABILITY	0.8427	0.4099	0.4957	0.3148
	Direct pulse	0.7198	0.4227	0.3692	0.3148
	Matched source state	0.7782	0.5157	0.4281	0.3704
	OPERATOR	**0.5704**	**0.6246**	**0.3623**	**0.4630**
	FULL L-STATE	0.7287	0.4102	0.3879	0.3519

I.4 FIXED EIGHT-STEP PULSES PREDICTING 16- AND 32-STEP RESPONSES

This separate post-hoc analysis asks whether a fixed short measurement retains predictive value at longer target horizons. For each original checkpoint, the four eight-step pulse measurements and current capabilities are passed through the original source-fitted readouts. Their eight-step total-response predictions are then evaluated unchanged against the observed total 16- and 32-step target responses: there is no multiplication by $h/8$, additional training, or target response used for calibration. Each target horizon counts from the original checkpoint, not from the end of a pulse branch. The fixed-input prediction uses only the eight-step measurements; the same-horizon pulse results serve as separate comparisons. Original 16-step branches and the completed 32-step extension supply the targets under the same scoring and training-order rules.

At 16 steps, the fixed-eight-step operator has RMSE 0.2943 versus capability's 0.4861, a 39.46% RMSE reduction (63.3% in MSE), with sign BA 0.7490 versus 0.3814. At 32 steps, RMSE is 0.3421 versus 0.4742, a 27.85% reduction (47.9% in MSE), with sign BA 0.7617 versus 0.4754. These results show that the original source calibration extracts predictive information from the fixed short measurement for target responses two and four times its per-pulse duration.

Using matched 16- and 32-step pulses instead gives operator RMSE 0.2878 and 0.3220. The fixed short measurement therefore retains much of the response-prediction benefit at both target horizons. This comparison reuses existing measurements and target outcomes; its post-hoc status is distinct from the prediction-before-target-training sequence of the matched-horizon extension.

J PROBE COST ACCOUNTING

A four-pulse measurement costs $4p$ optimizer steps and $16p$ training-example presentations per checkpoint. Matched 32/64/128-step measurements cost 128/256/512 optimizer steps. Running K target trials for H steps costs KH updates, giving the ratio $K_{\mathrm{BE}}(p, H) = 4p/H$; integer break-even counts are $\lceil 4p/H \rceil$.

At $p = 8$, the 32-step measurement equals two 16-step trials or one 32-step trial (Table 28). Two or four 32-step trials cost 64 or 128 steps; the shared measurement remains 32, with the 47.9% operator MSE reduction in Table 27. Target horizons count all steps from the original checkpoint.

Table 27: Post-hoc prediction of total 16- and 32-step target responses from fixed eight-step pulse measurements. All source-fitted predictions are reused unchanged across horizons. Units, scoring rules, and boldface follow Table 26; no target labels are used to refit or rescale predictions.

Steps	Readout	Raw RMSE ↓	Sign BA ↑	Regret ↓	Top-1 ↑
16	CAPABILITY	0.4861	0.3814	0.6063	0.3333
	Direct pulse	0.3111	0.5982	**0.3838**	**0.4074**
	Matched source state	0.3786	0.5787	0.4263	0.3519
	OPERATOR	**0.2943**	**0.7490**	0.4923	0.3704
	FULL L-STATE	0.3728	0.3567	0.4212	0.3704
32	CAPABILITY	0.4742	0.4754	0.5603	0.3889
	Direct pulse	**0.3249**	0.4851	**0.3131**	0.4259
	Matched source state	0.3817	0.5769	0.3728	**0.4444**
	OPERATOR	0.3421	**0.7617**	0.4105	0.3333
	FULL L-STATE	0.3775	0.3251	0.3555	0.4259

Table 28: **Per-checkpoint probe costs and target-horizon break-even.** Top: eight-step training setup. Bottom: fixed four-pulse measurement with eight steps per pulse, compared with longer target trials. Break-even entries are measurement-acquisition resource ratios; integer trial counts round up. Delivered-model parity additionally includes training the selected action.

Resource	Four-pulse probe	One target action	Break-even
Optimizer steps	32	8	4.0
Training examples	128	32	4.0
Evaluation examples	800	320	2.5

Target steps	Probe steps	One trial (steps)	Four trials (steps)	Acquisition step ratio
8	32	8	32	4.0
16	32	16	64	2.0
32	32	32	128	1.0

The comparison covers training updates to obtain predictions or trial outcomes. Source calibration, readout computation, and capability evaluation are separate costs; reference evaluation counts appear above. Executing a subsequently chosen adaptation is additional to the measurement cost. If exhaustive trials retain the winning model, probe-guided delivery costs $4p + H$ training steps versus KH for exhaustive delivery. Their step parity is $K_{\text{delivery}} = 1 + 4p/H$, with integer threshold $\lceil 1 + 4p/H \rceil$; strict savings require $K > 1 + 4p/H$. For $p = 8$ and $H = 8, 16, 32$, parity occurs at $K = 5, 3, 2$, respectively. At matched horizons $p = H$, four candidate trials cost $4H$, while probing followed by delivery costs $5H$. Evaluation adds 800 example forwards to a four-probe measurement and must also be priced for target screening, with shared baseline evaluations counted once. The measured wall-clock comparisons in Appendix K.1 include these costs. Reuse applies to alternative actions or utility queries branching from the same checkpoint; an updated checkpoint requires a new state measurement.

K ADDITIONAL COST AND GENERALIZATION EVIDENCE

K.1 FIXED-ENDPOINT AND EQUAL-TIME COMPARISONS

The measurement price is the additional expense over capability-only selection. Common model loading, input preparation, and capability evaluation are charged consistently to both policies. Four probes add 428.92 s on GLM and 398.11 s on Yi for measurement and readout; two probes add

218.45 s and 201.83 s. Selected-target training is a separate component, which can differ because the policies choose different actions. The full delivered-model costs in Table 29 include all three parts. In optimizer steps, four probes followed by one 32-step target use $4 \times 8 + 32 = 64$ steps, and two probes use $2 \times 8 + 32 = 48$, versus $4 \times 32 = 128$ for four complete target trials. These counts distinguish measurement acquisition from completed training decisions. Appendix K.6 gives the incremental-time decomposition. A measurement can be reused for different queries at the same unchanged checkpoint; each query retains its CPU readout cost, and subsequent model updates require a new state assessment.

Equal total-time caps answer the additional question of how much measurement can be afforded before training. Figure 4 retains all registered budgets and six policies, replaying real trajectories with measured component sums. Final scoring is a separately accounted reporting expense. At a 480-s cold-start cap, four probes return the original checkpoint on all Yi states; successive halving already yields mean gains of 5.063/5.122 under repeated/fresh continuation. At the 32-step endpoint, the four-probe utility difference from halving is -0.159 / -0.261 on GLM and $+0.664$ / -0.186 on Yi. The source-selected two-probe variant lowers total cold-start endpoint cost by 20.4–24.8% relative to four probes; on GLM it selects the same actions as capability only, with the same endpoint utility and about 218 additional seconds. The main comparison therefore prices the observed reduction in wrong training choices, while the budget curves specify when preparation constrains deployment. Their 32-step plateaus reflect the measured horizon, rather than a training limit of the capability baseline. Historical source-acquisition time remains unrecorded and is accounted separately (Appendix K.2).

K.2 Measured Components of Decision Cost

Tables below separate cold and resident-model (warm) replay. Loading includes the observed full-weight integrity validation, tokenizer/model/adapter initialization, selection-input encoding, and applicable CPU module/rule initialization. Baseline includes original-adapter restoration and selection evaluation. Training includes branch preparation, encoding, optimizer creation, and executed updates. Screening includes only the policy-required selection evaluations; CPU prediction/ranking is listed separately in milliseconds. Storage includes only policy-required saves and adapter/Adam/RNG recovery. Independent final scoring of the returned model and baseline is separate from decision cost. Extra diagnostics, unused intermediate checkpoints/evaluations, recovery-validation probes, and upstream state construction remain research-collection costs. Python process startup and dependency imports lie outside the worker clock. Physical card assignments and shared load are retained in the measurement records. GLM and Yi result files use separate local filesystems; policy comparisons within each family reuse measured components from the same filesystem, and checkpoint storage/recovery costs include this placement.

The decision comparison includes capability-only selection, source-calibrated prediction after two target steps per candidate (B5), direct selection after eight steps, and successive halving at steps 2 and 8. All policies share the initial capability measurement. All feature counts refit their source readouts. Physical GPU assignments may differ between measurement and target phases; all measurements use shared A6000 hardware.

The retained source experiment contains 81 states. Its main probe/target branches and state construction account for 5760 configured training steps. The additional 438 completed calibration branches account for 3672 configured training steps and 48 zero-gradient optimizer calls. The latter controls invoke the optimizer without a model forward pass. The two-step target observations used by the source-calibrated early baseline are prefixes of the retained target branches, with 0 additional training steps. These counts are configuration-derived and cross-checked against completed branches; historical measured GPU, training, evaluation and full-worker elapsed times remain unrecorded. The program-created attempt timestamp and completion timestamp recover the calendar windows below. Their sums describe overlapping worker windows, rather than active GPU time or a stopwatch total.

The sums of the main and calibration worker windows are 18919.3 and 8981.8 seconds, respectively (combined window sum 27901.1 seconds). These timestamp-window sums remain separate from all measured per-decision costs.

Interrupted attempts and collection overhead. Completed branch trajectories provide the fixed policy replay. Legal retries retain the original completed branches and generate new attempts for

Table 29: **Cost and utility of a delivered 32-step target model.** Cold-start mean component-composed seconds and independent final utility U for all six policies. Each family/setting uses nine states from three trajectories. This fixed-endpoint table uses the largest registered optimizer-step cap, under which every policy completes 32 target steps; actual optimizer steps differ across policies. Higher U is better. Figure 4 instead compares equal total-time caps. Required CPU decisions are included; independent final scoring is reported separately in Appendix K.2.

	GLM				Yi			
	Repeated		Fresh		Repeated		Fresh	
Method	Seconds	U	Seconds	U	Seconds	U	Seconds	U
Capability only	543.6	-8.905	546.3	-8.930	343.6	-10.801	347.0	-11.262
Target early 2	780.9	-9.184	775.4	-9.179	498.3	-14.107	495.1	-13.614
Halving 2–8–32	874.5	-8.587	889.3	-8.559	584.3	-10.747	581.2	-10.305
Four probes	976.8	-8.746	960.3	-8.819	741.0	-10.082	744.1	-10.490
Two probes	762.0	-8.905	764.7	-8.930	568.2	-10.490	559.9	-10.340
Direct early 8	887.9	-8.509	878.2	-8.485	581.6	-10.747	577.5	-10.305

Table 30: GLM: Repeated samples. Mean measured component-composed cost for the same 32-step returned targets. All columns use seconds except CPU (milliseconds). Final scoring is an additional reporting expense, including its input encoding exactly once.

Method	Start	Load	Baseline	Train	Screen	Store	CPU (ms)	Decision	Final
Capability only	cold	264.5	48.9	229.8	0.0	0.4	0.360	543.6	96.7
Target early 2	cold	264.5	48.9	286.7	178.8	2.0	0.345	780.9	100.1
Halving 2–8–32	cold	264.5	48.9	296.0	262.1	2.9	0.038	874.5	96.2
Four probes	cold	264.5	48.9	467.4	195.6	0.3	0.297	976.8	93.4
Two probes	cold	264.5	48.9	349.1	99.2	0.4	0.217	762.0	96.7
Direct early 8	cold	264.5	48.9	389.7	182.9	1.9	0.024	887.9	95.9
Capability only	warm	0.0	48.9	229.8	0.0	0.4	0.360	279.0	96.7
Target early 2	warm	0.0	48.9	286.7	178.8	2.0	0.345	516.3	100.1
Halving 2–8–32	warm	0.0	48.9	296.0	262.1	2.9	0.038	609.9	96.2
Four probes	warm	0.0	48.9	467.4	195.6	0.3	0.297	712.2	93.4
Two probes	warm	0.0	48.9	349.1	99.2	0.4	0.217	497.5	96.7
Direct early 8	warm	0.0	48.9	389.7	182.9	1.9	0.024	623.3	95.9

Table 31: GLM: Fresh continuation. Mean measured component-composed cost for the same 32-step returned targets. All columns use seconds except CPU (milliseconds). Final scoring is an additional reporting expense, including its input encoding exactly once.

Method	Start	Load	Baseline	Train	Screen	Store	CPU (ms)	Decision	Final
Capability only	cold	264.5	48.9	232.5	0.0	0.3	0.360	546.3	98.2
Target early 2	cold	264.5	48.9	271.8	188.2	1.9	0.341	775.4	100.8
Halving 2–8–32	cold	264.5	48.9	294.0	278.8	3.1	0.045	889.3	99.1
Four probes	cold	264.5	48.9	450.9	195.6	0.3	0.297	960.3	94.6
Two probes	cold	264.5	48.9	351.8	99.2	0.3	0.217	764.7	98.2
Direct early 8	cold	264.5	48.9	380.5	182.4	1.9	0.027	878.2	99.9
Capability only	warm	0.0	48.9	232.5	0.0	0.3	0.360	281.8	98.2
Target early 2	warm	0.0	48.9	271.8	188.2	1.9	0.341	510.8	100.8
Halving 2–8–32	warm	0.0	48.9	294.0	278.8	3.1	0.045	624.8	99.1
Four probes	warm	0.0	48.9	450.9	195.6	0.3	0.297	695.7	94.6
Two probes	warm	0.0	48.9	351.8	99.2	0.3	0.217	500.2	98.2
Direct early 8	warm	0.0	48.9	380.5	182.4	1.9	0.027	613.7	99.9

unfinished branches. Target workers must have identical encoded evaluation inputs and question order,

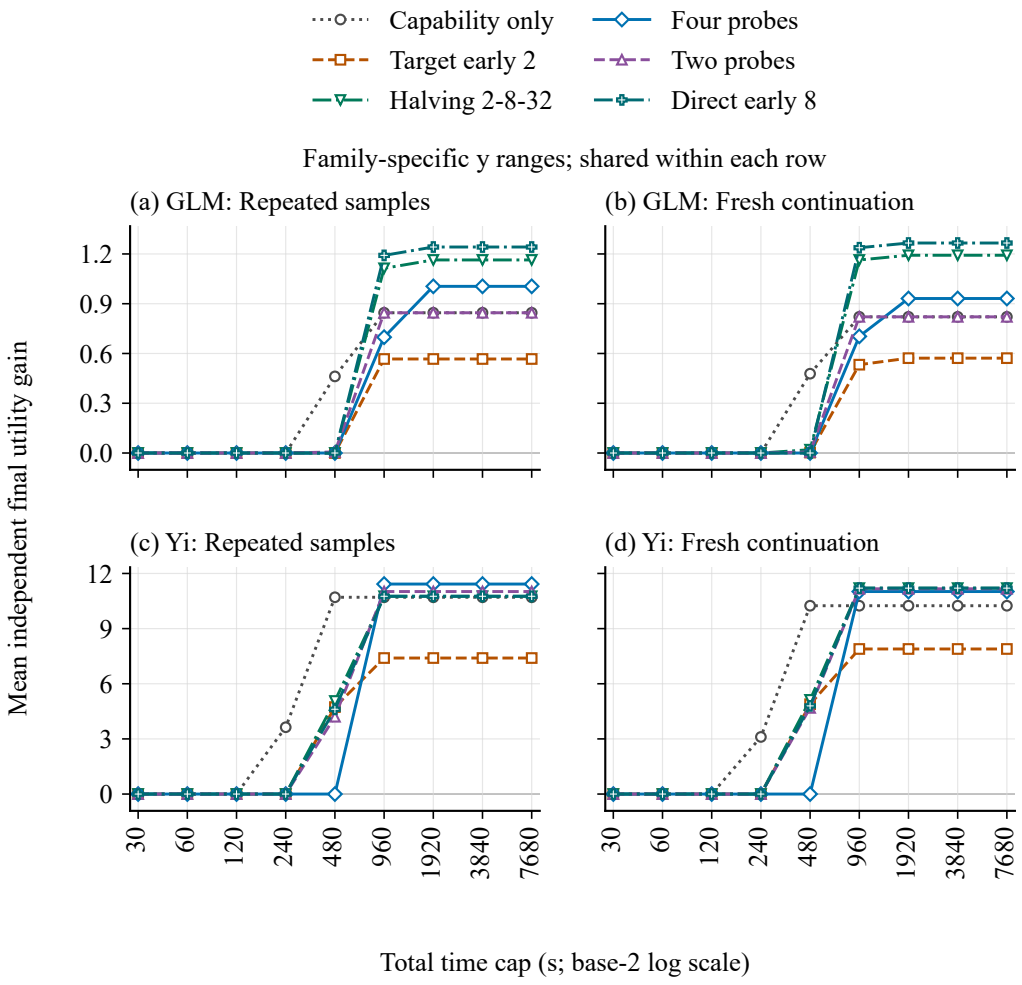

Figure 4: **Mean independent final utility gain under every registered total-time cap.** Rows are GLM and Yi; columns are repeated samples and fresh continuation. All six cold-start policies and all registered caps are retained, with family-specific vertical ranges shared within each row and a base-2 logarithmic time axis. Each point is the mean over nine states from three trajectories; lines only connect registered points. The plot shows means only; descriptive trajectory-level intervals for paired comparisons appear in Appendix K.5. Caps apply to decision costs composed from measured components, including the common baseline, required screening and CPU decisions. Final-reporting costs are accounted separately; final scores determine neither actions nor checkpoints. Original-checkpoint returns have zero gain.

consistent baseline vectors, and the same A6000 model. Final-reporting baseline and input-encoding costs always use the original fixed math-action worker, independently of timing and outcomes. A separate collection ledger enumerates every discovered branch-attempt and worker-shared timer event once, including failed and unfinished attempts. State-generation timers, when present, form a separate upstream category. Worker/process wall clocks contain these components and are listed separately. Collection-event sums describe elapsed component work across potentially parallel workers; they are neither GPU-busy time nor observed end-to-end policy latency, and are not added to each hypothetical policy budget. Attempt status identifies the enclosing recorded attempt, not the usefulness of all work within it. Shared preparation by a worker that later fails can also support completed branches. Failed-branch work and retry-worker preparation are therefore listed separately; the failed-worker category is not an estimate of entirely wasted work or of the causal incremental cost of failure. The discovered records contain 1 additional branch attempt and 1 additional worker attempt beyond the

Table 32: Yi: Repeated samples. Mean measured component-composed cost for the same 32-step returned targets. All columns use seconds except CPU (milliseconds). Final scoring is an additional reporting expense, including its input encoding exactly once.

Method	Start	Load	Baseline	Train	Screen	Store	CPU (ms)	Decision	Final
Capability only	cold	165.6	46.8	130.9	0.0	0.3	0.275	343.6	61.7
Target early 2	cold	165.6	46.8	169.4	114.9	1.6	0.324	498.3	62.6
Halving 2–8–32	cold	165.6	46.8	201.3	167.9	2.7	0.032	584.3	63.4
Four probes	cold	165.6	46.8	349.2	179.1	0.3	0.280	741.0	61.5
Two probes	cold	165.6	46.8	265.6	90.0	0.3	0.205	568.2	64.5
Direct early 8	cold	165.6	46.8	255.0	112.6	1.6	0.023	581.6	63.4
Capability only	warm	0.0	46.8	130.9	0.0	0.3	0.275	178.0	61.7
Target early 2	warm	0.0	46.8	169.4	114.9	1.6	0.324	332.7	62.6
Halving 2–8–32	warm	0.0	46.8	201.3	167.9	2.7	0.032	418.7	63.4
Four probes	warm	0.0	46.8	349.2	179.1	0.3	0.280	575.4	61.5
Two probes	warm	0.0	46.8	265.6	90.0	0.3	0.205	402.6	64.5
Direct early 8	warm	0.0	46.8	255.0	112.6	1.6	0.023	416.0	63.4

Table 33: Yi: Fresh continuation. Mean measured component-composed cost for the same 32-step returned targets. All columns use seconds except CPU (milliseconds). Final scoring is an additional reporting expense, including its input encoding exactly once.

Method	Start	Load	Baseline	Train	Screen	Store	CPU (ms)	Decision	Final
Capability only	cold	165.6	46.8	134.4	0.0	0.3	0.275	347.0	62.2
Target early 2	cold	165.6	46.8	168.3	112.8	1.6	0.482	495.1	62.3
Halving 2–8–32	cold	165.6	46.8	198.5	167.7	2.6	0.040	581.2	62.7
Four probes	cold	165.6	46.8	352.4	179.1	0.3	0.280	744.1	62.2
Two probes	cold	165.6	46.8	257.3	90.0	0.2	0.205	559.9	61.1
Direct early 8	cold	165.6	46.8	252.9	110.7	1.6	0.025	577.5	62.7
Capability only	warm	0.0	46.8	134.4	0.0	0.3	0.275	181.4	62.2
Target early 2	warm	0.0	46.8	168.3	112.8	1.6	0.482	329.5	62.3
Halving 2–8–32	warm	0.0	46.8	198.5	167.7	2.6	0.040	415.6	62.7
Four probes	warm	0.0	46.8	352.4	179.1	0.3	0.280	578.5	62.2
Two probes	warm	0.0	46.8	257.3	90.0	0.2	0.205	394.3	61.1
Direct early 8	warm	0.0	46.8	252.9	110.7	1.6	0.025	411.9	62.7

Table 34: One-time source costs, separate from per-decision target costs. CPU fitting was measured once for the unchanged frozen readouts, including grouped source-trajectory validation. Historical acquisition time remains unrecorded.

Component	Seconds
Historical source-response collection	Unrecorded
Source records: read and hash	0.003892
Source fitting and serialization verification	1.082612

first attempt for each branch or worker. These counts describe retained attempts; the component table retains their measured expenses without allocating the same retry expense to every policy.

K.3 Source-only Probe Selection

Candidate prefixes follow schema mapping, rule chaining, symbol rewriting, and table aggregation. Each leave-one-family-out model independently fits its standardization, grouped ridge selection, and prescribed rank-deficiency repair. The selected two-probe readout uses schema mapping and rule chaining. Selection minimizes equal-family mean normalized action regret across the original

Table 35: Retained source experiment: observed completed branch counts and configuration-derived optimizer calls. Explicit calibration repeats are planned controls. Counts do not reconstruct discarded, overwritten or unrecorded attempts.

Component	Completed branches	Training steps	Zero-gradient calls
Main probes	324	2592	0
Main targets	324	2592	0
State construction	72	576	0
Calibration: reliability repeats	144	1152	0
Calibration: controlled mixtures	108	864	0
Calibration: step scan	108	1080	0
Calibration: amplitude controls	78	576	48

Table 36: Historical program-timestamp windows in UTC, reconstructed from attempt identifiers and completion records. Window length is an artifact-timestamp difference; it is not a measured runtime, GPU duration, or isolated training/evaluation cost. Parallel windows may overlap. Timestamps are displayed to the second.

Family	Phase	Attempt created (UTC)	Completion (UTC)	Window (s)
Qwen	Main	2026-08-29 08:56:11	2026-08-29 10:30:05	5633.6
Qwen	Calibration	2026-08-29 10:33:33	2026-08-29 11:21:45	2891.7
Mistral	Main	2026-08-29 09:43:50	2026-08-29 11:30:37	6406.9
Mistral	Calibration	2026-08-29 11:32:40	2026-08-29 12:19:37	2817.5
OLMo	Main	2026-08-29 09:44:47	2026-08-29 11:39:26	6878.8
OLMo	Calibration	2026-08-29 11:41:43	2026-08-29 12:36:16	3272.6

Table 37: Discovered experiment-collection timer events, including unsuccessful attempts. Scope separates branch work, shared worker preparation, and upstream state generation. Each event is counted once; containing wall clocks are excluded from these sums. Historical GLM upstream generation remains unrecorded.

Family	Scope	Attempt status	Events	Component seconds
GLM	Branch attempts	completed	7416	46579.9
GLM	Shared worker	completed	66	2772.9
Yi	Branch attempts	completed	7416	30855.1
Yi	Branch attempts	failed	31	340.8
Yi	State generation	completed	63	2144.4
Yi	Shared worker	completed	63	2009.3
Yi	Shared worker	failed	10	418.7

six utilities among the three reduced candidates, with fewer probe steps and then lower RMSE as tie-breakers. Four probes provide the full comparator. These development results describe the source choice; the reduced readout's generalization is measured by the external family/setting experiment. The reduced readout is an empirical feature-reduction variant; the four-column operator conditions apply to the complete operator.

K.4 Response Prediction Under the Crossed Settings

Source-frozen eight-step response predictions are evaluated unchanged at 16 and 32 target steps. RMSE uses raw capability-change units. Direction balanced accuracy first averages available pos-

Table 38: Source-only development comparison across Qwen, Mistral, and OLMo. Probe steps are configuration-derived. Regret and action top-1 evaluate selected training actions; RMSE is source-standardized. Direction BA evaluates response signs. The independently chosen two-probe variant retains its higher error and regret relative to the full comparator.

Probes	Probe steps	Regret	Action top-1	RMSE	Direction BA
1	8	0.2996	0.4074	9.1078	0.6039
2 (selected)	16	0.2593	0.4938	9.6145	0.6150
3	24	0.2621	0.5226	8.3583	0.5979
4	32	0.2477	0.5412	8.2925	0.5885

itive/negative recall within each capability coordinate and then across determinate coordinates, retaining the source-frozen noise dead zones. Brackets report the 2.5th and 97.5th percentiles from exact whole-trajectory resampling; three trajectories produce 27 ordered resamples. They summarize this experiment descriptively. Selection-set and independent final-set responses appear in separate tables; only selection-set observations enter decisions. The original duration-study evaluation set corresponds to the selection set here. Undefined sign metrics are shown by a dash.

K.5 All Registered Time-budget Comparisons

Every registered cold-start time cap is retained below for the two probe policies against the source-calibrated two-step target baseline (B5) and successive halving (SH). Entries are mean paired differences in independent final balanced utility, followed by descriptive 95% whole-trajectory resampling intervals. Positive values favor the probe policy. The same nine states and total-time cap are paired within each family and setting. The three trajectories, rather than their related checkpoints, are resampled (5,000 resamples with the fixed analysis seed). Incomplete measurement can return the original checkpoint while incurring measurement costs. Both startup conditions and every registered budget remain in the accompanying result tables.

A retained adverse case. In Yi fresh continuation, the largest observed deficit of four probes against SH at the registered 1,920-s cap occurred for trajectory seed 20260901, state 0. Four probes selected science and SH selected reading; both returned 32-step targets, and the paired independent final utility difference was -2.3768. This case is reported retrospectively and did not affect the frozen selection rules. The 3,840-s and 7,680-s caps return the same terminal checkpoints; the repeated plateau values are not additional independent evidence.

K.6 Capability-Referenced Value and Matched Readout Controls

The capability-referenced analysis uses all original states, six objectives, registered budgets, and cold/warm conditions. The main table reports the pre-existing balanced objective. Table 47 retains the mean utility differences for every objective at the common 32-step endpoint. Utility is evaluated on an independent final measurement set after the source-frozen action has been chosen.

Incremental measurement versus target-training cost. The capability policy's shared cold-start preparation, including its CPU readout, averages 313.42 s for GLM and 212.36 s for Yi. Subtracting this same preparation from the probe policy isolates its extra measurement and readout. The full difference in delivered-model expense also includes the time difference between the selected target actions (Table 48). Shared preparation and the extra measurement are charged once for a single decision. If a fixed checkpoint supports several genuine prediction or selection queries, the acquired probe responses can be shared, while each additional CPU query is charged separately. This is a reuse scenario, not a measured multi-query deployment. Continued training changes the state. Independent final scoring, historical source acquisition, and research collection remain separately accounted expenses under the earlier tables' definitions.

Matched finite-class readout comparison. We additionally compare two inputs, current capability and capability plus four probes, using the same 30 candidate configurations. Linear and quadratic

Table 39: GLM: independent final-set response metrics for every readout and both settings. Each method/setting/horizon uses nine states, four actions per state, and five response coordinates. Brackets are descriptive trajectory-resampling intervals.

		16 steps		32 steps	
Setting	Readout	Raw RMSE	Direction BA	Raw RMSE	Direction BA
Repeated	Capability only	0.471 [0.465, 0.478]	0.439 [0.404, 0.465]	0.452 [0.449, 0.456]	0.429 [0.376, 0.609]
Repeated	Target early 2	0.272 [0.269, 0.276]	0.775 [0.756, 0.806]	0.276 [0.274, 0.279]	0.691 [0.656, 0.841]
Repeated	Calibrated operator	0.268 [0.267, 0.268]	0.810 [0.797, 0.847]	0.326 [0.324, 0.327]	0.654 [0.574, 0.743]
Repeated	Four probes	0.353 [0.348, 0.358]	0.443 [0.414, 0.465]	0.354 [0.350, 0.357]	0.390 [0.362, 0.551]
Repeated	Two probes	0.332 [0.329, 0.335]	0.443 [0.414, 0.465]	0.334 [0.331, 0.336]	0.415 [0.365, 0.576]
Fresh	Capability only	0.473 [0.470, 0.477]	0.464 [0.460, 0.504]	0.460 [0.456, 0.464]	0.469 [0.436, 0.496]
Fresh	Target early 2	0.275 [0.272, 0.279]	0.789 [0.777, 0.797]	0.280 [0.279, 0.281]	0.775 [0.770, 0.779]
Fresh	Calibrated operator	0.269 [0.269, 0.269]	0.772 [0.747, 0.796]	0.325 [0.324, 0.326]	0.719 [0.687, 0.745]
Fresh	Four probes	0.356 [0.353, 0.360]	0.422 [0.404, 0.481]	0.361 [0.358, 0.365]	0.411 [0.406, 0.416]
Fresh	Two probes	0.333 [0.331, 0.336]	0.447 [0.406, 0.511]	0.338 [0.335, 0.341]	0.457 [0.422, 0.478]

Table 40: Yi: independent final-set response metrics for every readout and both settings. Each method/setting/horizon uses nine states, four actions per state, and five response coordinates. Brackets are descriptive trajectory-resampling intervals.

		16 steps		32 steps	
Setting	Readout	Raw RMSE	Direction BA	Raw RMSE	Direction BA
Repeated	Capability only	4.072 [4.062, 4.084]	0.702 [0.693, 0.729]	4.241 [4.227, 4.254]	0.687 [0.673, 0.700]
Repeated	Target early 2	2.439 [2.430, 2.446]	0.645 [0.596, 0.710]	3.032 [3.005, 3.066]	0.622 [0.596, 0.655]
Repeated	Calibrated operator	3.061 [3.050, 3.075]	0.695 [0.667, 0.746]	3.654 [3.636, 3.673]	0.704 [0.684, 0.726]
Repeated	Four probes	2.791 [2.774, 2.808]	0.733 [0.701, 0.769]	3.368 [3.355, 3.383]	0.710 [0.701, 0.725]
Repeated	Two probes	3.178 [3.172, 3.183]	0.706 [0.684, 0.740]	3.281 [3.271, 3.293]	0.697 [0.688, 0.708]
Fresh	Capability only	4.083 [4.070, 4.094]	0.728 [0.694, 0.787]	4.318 [4.308, 4.328]	0.671 [0.662, 0.685]
Fresh	Target early 2	2.428 [2.423, 2.433]	0.651 [0.600, 0.722]	3.044 [3.008, 3.078]	0.614 [0.593, 0.645]
Fresh	Calibrated operator	3.045 [3.036, 3.054]	0.699 [0.670, 0.745]	3.664 [3.635, 3.700]	0.705 [0.690, 0.719]
Fresh	Four probes	2.778 [2.763, 2.792]	0.737 [0.716, 0.781]	3.387 [3.358, 3.422]	0.686 [0.682, 0.694]
Fresh	Two probes	3.162 [3.157, 3.166]	0.713 [0.712, 0.741]	3.292 [3.241, 3.345]	0.674 [0.665, 0.686]

predictors use ridge penalties in $\{0.01, 0.1, 1, 10, 100\}$; radial-basis kernel predictors combine these penalties with kernel parameters $\{0.01, 0.1, 1, 10\}$. Training-fold input/output standardization and

Table 41: GLM: selection-set response metrics for every readout and both settings. Each method/setting/horizon uses nine states, four actions per state, and five response coordinates. Brackets are descriptive trajectory-resampling intervals.

		16 steps		32 steps	
Setting	Readout	Raw RMSE	Direction BA	Raw RMSE	Direction BA
Repeated	Capability only	0.486 [0.484, 0.488]	0.381 [0.364, 0.397]	0.474 [0.469, 0.480]	0.475 [0.455, 0.508]
Repeated	Target early 2	0.290 [0.284, 0.295]	0.581 [0.571, 0.722]	0.303 [0.302, 0.305]	0.610 [0.535, 0.678]
Repeated	Calibrated operator	0.294 [0.293, 0.296]	0.749 [0.733, 0.772]	0.342 [0.341, 0.343]	0.762 [0.727, 0.801]
Repeated	Four probes	0.373 [0.372, 0.374]	0.357 [0.344, 0.367]	0.377 [0.374, 0.381]	0.325 [0.270, 0.356]
Repeated	Two probes	0.347 [0.347, 0.348]	0.357 [0.344, 0.367]	0.349 [0.346, 0.352]	0.345 [0.270, 0.398]
Fresh	Capability only	0.490 [0.487, 0.494]	0.392 [0.370, 0.415]	0.492 [0.488, 0.495]	0.481 [0.328, 0.527]
Fresh	Target early 2	0.293 [0.289, 0.297]	0.748 [0.727, 0.775]	0.317 [0.313, 0.322]	0.638 [0.582, 0.697]
Fresh	Calibrated operator	0.298 [0.297, 0.298]	0.742 [0.730, 0.756]	0.357 [0.355, 0.360]	0.656 [0.643, 0.771]
Fresh	Four probes	0.376 [0.373, 0.381]	0.371 [0.370, 0.372]	0.397 [0.393, 0.399]	0.336 [0.294, 0.353]
Fresh	Two probes	0.352 [0.350, 0.354]	0.396 [0.370, 0.424]	0.367 [0.364, 0.371]	0.386 [0.312, 0.425]

Table 42: Yi: selection-set response metrics for every readout and both settings. Each method/setting/horizon uses nine states, four actions per state, and five response coordinates. Brackets are descriptive trajectory-resampling intervals.

		16 steps		32 steps	
Setting	Readout	Raw RMSE	Direction BA	Raw RMSE	Direction BA
Repeated	Capability only	4.119 [4.107, 4.134]	0.725 [0.701, 0.769]	4.295 [4.273, 4.318]	0.684 [0.678, 0.689]
Repeated	Target early 2	2.470 [2.463, 2.477]	0.661 [0.601, 0.751]	3.048 [3.018, 3.087]	0.622 [0.596, 0.656]
Repeated	Calibrated operator	3.101 [3.093, 3.110]	0.701 [0.672, 0.747]	3.666 [3.647, 3.688]	0.701 [0.684, 0.720]
Repeated	Four probes	2.830 [2.816, 2.844]	0.750 [0.726, 0.795]	3.383 [3.368, 3.402]	0.706 [0.700, 0.719]
Repeated	Two probes	3.211 [3.204, 3.217]	0.718 [0.713, 0.744]	3.290 [3.277, 3.304]	0.693 [0.683, 0.706]
Fresh	Capability only	4.136 [4.128, 4.145]	0.745 [0.704, 0.800]	4.374 [4.363, 4.390]	0.671 [0.661, 0.683]
Fresh	Target early 2	2.462 [2.457, 2.468]	0.657 [0.601, 0.745]	3.065 [3.025, 3.101]	0.618 [0.596, 0.643]
Fresh	Calibrated operator	3.084 [3.078, 3.089]	0.701 [0.672, 0.747]	3.684 [3.651, 3.720]	0.699 [0.686, 0.718]
Fresh	Four probes	2.819 [2.805, 2.829]	0.746 [0.718, 0.803]	3.408 [3.376, 3.445]	0.692 [0.687, 0.699]
Fresh	Two probes	3.198 [3.192, 3.206]	0.719 [0.714, 0.744]	3.314 [3.267, 3.363]	0.681 [0.674, 0.689]

an intercept are applied consistently; quadratic features include all pairwise products. Each input separately selects a configuration by leave-one-family-out validation on Qwen, Mistral, and OLMo.

Table 43: GLM: repeated samples. Cold-start paired utility differences at all nine registered time caps. Each pair uses nine states from three trajectories; brackets are descriptive trajectory-level intervals. Unfavorable differences are retained. B5 denotes source-calibrated two-step screening; SH denotes successive halving.

Time cap (s)	Four probes minus B5	Four probes minus SH	Two probes minus B5	Two probes minus SH
30	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
60	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
120	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
240	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
480	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.007$ $[+0.000, +0.011]$	$+0.007$ $[+0.000, +0.011]$
960	$+0.131$ $[+0.006, +0.314]$	-0.415 $[-0.576, -0.161]$	$+0.279$ $[+0.262, +0.303]$	-0.268 $[-0.388, -0.173]$
1920	$+0.438$ $[+0.416, +0.480]$	-0.159 $[-0.306, +0.005]$	$+0.279$ $[+0.262, +0.303]$	-0.318 $[-0.461, -0.173]$
3840	$+0.438$ $[+0.416, +0.480]$	-0.159 $[-0.306, +0.005]$	$+0.279$ $[+0.262, +0.303]$	-0.318 $[-0.461, -0.173]$
7680	$+0.438$ $[+0.416, +0.480]$	-0.159 $[-0.306, +0.005]$	$+0.279$ $[+0.262, +0.303]$	-0.318 $[-0.461, -0.173]$

Table 44: GLM: fresh continuation. Cold-start paired utility differences at all nine registered time caps. Each pair uses nine states from three trajectories; brackets are descriptive trajectory-level intervals. Unfavorable differences are retained. B5 denotes source-calibrated two-step screening; SH denotes successive halving.

Time cap (s)	Four probes minus B5	Four probes minus SH	Two probes minus B5	Two probes minus SH
30	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
60	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
120	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
240	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
480	-0.003 $[-0.008, +0.000]$	-0.020 $[-0.060, +0.000]$	$+0.004$ $[+0.000, +0.009]$	-0.014 $[-0.049, +0.009]$
960	$+0.171$ $[+0.078, +0.260]$	-0.460 $[-0.630, -0.242]$	$+0.288$ $[+0.256, +0.350]$	-0.343 $[-0.448, -0.246]$
1920	$+0.360$ $[+0.317, +0.395]$	-0.261 $[-0.382, -0.108]$	$+0.249$ $[+0.231, +0.259]$	-0.371 $[-0.490, -0.246]$
3840	$+0.360$ $[+0.317, +0.395]$	-0.261 $[-0.382, -0.108]$	$+0.249$ $[+0.231, +0.259]$	-0.371 $[-0.490, -0.246]$
7680	$+0.360$ $[+0.317, +0.395]$	-0.261 $[-0.382, -0.108]$	$+0.249$ $[+0.231, +0.259]$	-0.371 $[-0.490, -0.246]$

One protocol minimizes equal-family standardized response MSE; the other minimizes equal-family normalized action regret averaged over the six existing objectives. Both selected configurations are refit on the 81 source records before external evaluation. Source labels retain the original eight-step horizon, and the resulting response predictions are evaluated at 32 steps without horizon rescaling.

Table 45: Yi: repeated samples. Cold-start paired utility differences at all nine registered time caps. Each pair uses nine states from three trajectories; brackets are descriptive trajectory-level intervals. Unfavorable differences are retained. B5 denotes source-calibrated two-step screening; SH denotes successive halving.

Time cap (s)	Four probes minus B5	Four probes minus SH	Two probes minus B5	Two probes minus SH
30	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
60	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
120	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
240	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
480	-4.732 $[-5.021, -4.164]$	-5.063 $[-5.224, -4.900]$	-0.527 $[-1.278, -0.052]$	-0.857 $[-2.014, -0.105]$
960	$+4.024$ $[+3.902, +4.121]$	$+0.664$ $[+0.467, +1.034]$	$+3.617$ $[+3.510, +3.791]$	$+0.257$ $[-0.007, +0.703]$
1920	$+4.024$ $[+3.902, +4.121]$	$+0.664$ $[+0.467, +1.034]$	$+3.617$ $[+3.510, +3.791]$	$+0.257$ $[-0.007, +0.703]$
3840	$+4.024$ $[+3.902, +4.121]$	$+0.664$ $[+0.467, +1.034]$	$+3.617$ $[+3.510, +3.791]$	$+0.257$ $[-0.007, +0.703]$
7680	$+4.024$ $[+3.902, +4.121]$	$+0.664$ $[+0.467, +1.034]$	$+3.617$ $[+3.510, +3.791]$	$+0.257$ $[-0.007, +0.703]$

Table 46: Yi: fresh continuation. Cold-start paired utility differences at all nine registered time caps. Each pair uses nine states from three trajectories; brackets are descriptive trajectory-level intervals. Unfavorable differences are retained. B5 denotes source-calibrated two-step screening; SH denotes successive halving.

Time cap (s)	Four probes minus B5	Four probes minus SH	Two probes minus B5	Two probes minus SH
30	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
60	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
120	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
240	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$	$+0.000$ $[+0.000, +0.000]$
480	-4.887 $[-5.263, -4.327]$	-5.122 $[-5.270, -5.005]$	-0.200 $[-1.441, +1.171]$	-0.435 $[-2.119, +1.343]$
960	$+3.124$ $[+2.855, +3.377]$	-0.186 $[-0.410, +0.155]$	$+3.274$ $[+3.004, +3.578]$	-0.036 $[-0.548, +0.421]$
1920	$+3.124$ $[+2.855, +3.377]$	-0.186 $[-0.410, +0.155]$	$+3.274$ $[+3.004, +3.578]$	-0.036 $[-0.548, +0.421]$
3840	$+3.124$ $[+2.855, +3.377]$	-0.186 $[-0.410, +0.155]$	$+3.274$ $[+3.004, +3.578]$	-0.036 $[-0.548, +0.421]$
7680	$+3.124$ $[+2.855, +3.377]$	-0.186 $[-0.410, +0.155]$	$+3.274$ $[+3.004, +3.578]$	-0.036 $[-0.548, +0.421]$

GLM/Yi outcomes enter only the subsequent scoring. These are additional post-hoc controls; the original readouts and action choices remain fixed.

Table 47: Full four-probe L-STATE minus the original capability-only baseline: mean independent final utility difference at 32 target steps. All six existing objectives are retained. Each cell uses nine states from three trajectories. Objective scales differ, so differences are interpreted within columns.

Family	Setting	Balanced	Math	Code	Science	Reading	General
GLM	Repeated	+0.1592	+0.2085	+0.0000	+0.0000	+0.0026	+0.0211
GLM	Fresh	+0.1105	+0.2603	+0.0000	+0.0000	+0.0041	+0.0117
Yi	Repeated	+0.7188	+0.2579	+0.2981	+0.0000	+1.4361	+0.0000
Yi	Fresh	+0.7714	+0.3253	+0.3012	+0.0000	+1.2145	+0.0000

Table 48: Incremental component-composed seconds for four probes relative to capability-only selection. Measurement includes additional probe/readout costs; the target term is the cost difference between the two selected 32-step training actions. The two terms sum to the total difference. Shared cold-start preparation cancels. The same nine states are paired in each row.

Family	Setting	Measurement	Target difference	Total difference
GLM	Repeated	428.92	+4.28	433.20
GLM	Fresh	428.92	-14.97	413.96
Yi	Repeated	398.11	-0.71	397.40
Yi	Fresh	398.11	-0.97	397.14

Table 49: Adding four probes under matched readout search: mean 32-step balanced utility difference between capability-plus-probes and capability-only inputs. Configurations are selected on source families using either response error or decision regret. Each external cell contains nine states from three trajectories.

Family	Setting	Response-selected	Decision-selected
GLM	Repeated	+0.4379	+0.2878
GLM	Fresh	+0.3596	+0.2136
Yi	Repeated	+0.1875	+0.5313
Yi	Fresh	+0.0846	+0.6869

Capability-only response/decision selection chooses radial-basis kernels with penalty 0.01 and kernel parameter 1/0.01. Probe response selection chooses linear ridge with penalty 100; probe decision selection chooses the radial-basis kernel with penalty 0.01 and parameter 0.01. The paired mean decision differences in Table 49 are positive in all four cells under either selection protocol. Nonlinear capability readouts also improve some prediction errors: the decision-selected capability predictor reaches response RMSE 3.286/3.297 on Yi repeated/fresh continuation, below the original full-state readout's 3.368/3.387, while its balanced final utility is lower. The newly selected probe readouts on Yi also have lower balanced utility than the original frozen full-state version. Thus the empirical result is a comparison of specified source-fitted predictors and decisions. Theorem 3.1 addresses a different, conditional question: when two states have the same capability vector but different expected training responses, an arbitrary capability-only predictor faces an identification error. The present finite-class results neither measure that universal optimum nor turn close capability vectors into exact level-set counterexamples.

K.7 STATE-LEVEL INTERPRETATION OF THE CAPABILITY COMPARISON

We first compare the information supplied by four probes with the original capability-only baseline at a common 32-step target endpoint. GLM and Yi each contribute three trajectories with three related states; both repeated- and fresh-example continuations share their first eight steps. The full-state readout and action rule remain source-frozen. Independent final measurements show 36.9–38.8% lower response MSE and 24.8–68.8% lower action-selection regret across all four family/setting cells (Table 2). For a state s, regret is $\max_{a \in \mathcal{A}} U_{32}(s, a) - U_{32}(s, \widehat{a})$, using the same four candidate actions.

The percentage reduction compares mean regret within each family and setting. The retrospectively best action is used only for scoring; the decision itself uses no final outcomes. Utility weights $(1, 1, 1, 1, 0.5)$ combine five negative clipped-token-NLL coordinates.

Each cell contains three improved and six unchanged decisions, with no lower balanced final utility among these nine states. The improved decisions occur at GLM's initial states, where full L-STATE selects code instead of mathematics, and at Yi's later observed states, where it selects code instead of science. These observed corrections explain the value of response measurement beyond the static score vector. The paired intervals resample whole trajectories and summarize these three trajectories descriptively. Appendix K.6 retains all six utility objectives and examines richer capability-only readouts using source-selected linear, quadratic, and radial-basis predictors. In both matched selection protocols, adding probes improves mean balanced decision utility in all four cells. These finite-readout comparisons complement the conditional information obstruction in Theorem 3.1; they do not numerically identify the optimum over every possible capability-only predictor.